

**Eine epistemische Alternative:
Bayesianische Statistik und ihr Potential für die
empirische Pflegeforschung**

Masterarbeit zum Erlangen des akademischen Grades
„Master of Science“

im Masterstudiengang Pflegewissenschaft

an der pflegewissenschaftlichen Fakultät der
Philosophisch-Theologischen Hochschule Vallendar

vorgelegt von:

Michael Knop

Matrikelnummer 162234788005

Teutoburger Straße 67

33607 Bielefeld

Erstprüfer: **Prof. Dr. Albert Brühl**

Zweitprüferin: **Prof.in Dr.in Sandra Bensch**

vorgelegt am 08.08.2018

Das, wobei unsere Berechnungen versagen, nennen wir Zufall.

[Albert Einstein]

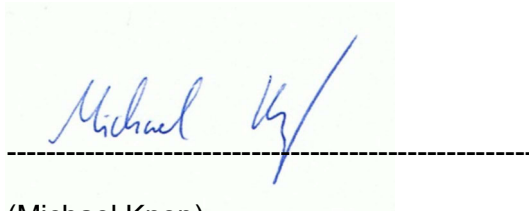
Inhaltsverzeichnis

Eigenständigkeitserklärung.....	1
Abkürzungsverzeichnis.....	2
Tabellen- und Abbildungsverzeichnis	4
1.Einleitung	6
2.Literaturrecherche und Methodik	8
3.Forschungsstand	9
4.Historische Betrachtung: Die Bayesianische Idee.....	11
4.1 Entwicklung vom 18. bis zum 19. Jahrhundert	12
4.2 Entwicklung ab dem 19. Jahrhundert	13
5. Mathematische Grundlagen und Operationalisierung	17
5.1 Grundsatz.....	18
5.1.1 <i>Priori</i> -Verteilung.....	19
5.1.2 Wahrscheinlichkeitsverteilungen.....	23
5.1.3 Likelihood	35
5.1.4 Posteriori-Verteilung.....	36
5.2 Voraussetzungen.....	40
5.3 Die Markow Chain Monte Carlo Methode.....	44
5.3.1 Metropolis-Hastings Algorithmus	45
5.3.2 Gibbs Sampling.....	47
5.4 Bayesianische Verfahren	49
5.4.1 Bayes-Faktor	49
5.4.2 Bayesianischer Hypothesentest.....	51
5.4.3 Bayesianische Lineare Modelle.....	57
5.4.4 Bayesianische Mehrebenenanalyse.....	61
5.4.5 Anwendungsbeispiel: Bayes-Faktor und bayesianisches Hypothesentesten.....	64
6.Disziplinbezogene Verortung.....	73
6.1 Erkenntnis und die Definition von Wahrscheinlichkeit.....	73
6.2 Empirische Pflegeforschung und Statistik.....	75
6.3 Professionalisierung der Pflege.....	76

7. Diskussion und Limitation	79
8. Zusammenfassung und Ausblick	82
Literaturverzeichnis	85
Einverständniserklärung	89
Anhang: Rechercheprotokoll	90

Eigenständigkeitserklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbstständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internetquellen – benutzt habe und die Arbeit von mir vorher nicht in einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD).



(Michael Knop)

Bielefeld, den 08.08.2018

Abkürzungsverzeichnis

Abb.	- Abbildung
al.	- alii / aliae
ALM	- Allgemeine Lineare Modelle
ANOVA	- Analysis of variance
bspw.	- Beispielsweise
Diagr.	- Diagramm
EAP	- expected a posteriori
ebd.	- ebenda
EBM	- Evidence Based Medicine
EBN	- Evidence Based Nursing
engl.	- englisch
f.	- folgende
ff.	- fortfolgende
GLM	- Generalisierte Lineare Modelle
HDI	- highest posterior density interval
HPD	- highest posterior density
i.i.d.	- independent and identically distributed
JB1	- Joanna Briggs Institute
MCMC	- Markow Chain Monte Carlo
MEA	- Mehrebenenanalyse
MLA	- Multilevel Anaysis
n.	- nach
NHST	- null hypothesis significance testing
PRISMA	- Preferred Reporting Items for Systematic Reviews and Meta-Analyses
s.	- siehe

S.	- Seite
sog.	- Sogenannte/n/r
Tab.	- Tabelle
vgl.	- Vergleiche
z.B.	- zum Beispiel

Tabellen- und Abbildungsverzeichnis

Bezeichnung	Inhalt	Seite
Tabelle 1	Zusammenfassung aller pflegerelevanten Artikel mit Verwendung bayesianischer Verfahren	10
Tabelle 2	Jeffreys' prior bei unterschiedlicher Likelihood	21
Tabelle 3	Zusammenfassung konjugierter priors	34
Tabelle 4	Hilfe zur Interpretation des Bayes-Faktor	50
Tabelle 5	Kennzahlen der Studie von Dauer et al.(2006)	65
Tabelle 6	Kennwerte der T-Statistik aus der Studie von Dauer et al. (2006)	65
Diagramm 1	Verschiedene priors und ihre Zusammenhänge	22
Abbildung 1	Graph einer stetigen Gleichverteilung	20
Abbildung 2	Dichtefunktion einer Normalverteilung	24
Abbildung 3	Geometrische Verteilung	26
Abbildung 4	Binomialverteilung	27
Abbildung 5	Beta-Verteilung	28
Abbildung 6	Verteilung einer zweidimensionalen Copula	30
Abbildung 7	Poisson-Verteilung	31
Abbildung 8	Gamma-Verteilung	32
Abbildung 9	Cauchy-Verteilung	24
Abbildung 10	Annähernd gleichverteilte Beta-Funktion als prior	38
Abbildung 11	Annähernd zentralisierte Beta-Funktion als prior	39
Abbildung 12	Rechtsseitige Beta-Funktion als prior	39
Abbildung 13	Beziehung bayesianischer Voraussetzungen	43
Abbildung 14	Schematischer Ablauf des Metropolis-Hastings Algorithmus	46

Bezeichnung	Inhalt	Seite
Abbildung 15	Hierarchisches Diagramm für bayesianische Parameterschätzung	55
Abbildung 16	Posteriori-Verteilungen unterschiedlicher Kenngrößen	56
Abbildung 17	Posteriori-Verteilungen zweier Regressionskoeffizienten	60
Abbildung 18	Hierarchisches Modell zur Bestimmung der Heterogenität zwischen den Gruppen einer mehrdimensionalen bayesianischen GLR	62
Abbildung 19	Posteriori-Verteilung eines slope Parameters	63
Abbildung 20	Histogramm-artige Abbildungen der Testergebnisse aus der Studie von Dauer et al. (2006)	68
Abbildung 21	Ergebnisse der MCMC-gestützten Approximation von Posteriori-Verteilungen unterschiedlicher Parameter	70

1. Einleitung

Im Vergleich zu anderen wissenschaftlichen Disziplinen befindet sich die Pflegewissenschaft (insbesondere in Deutschland) in einer relativ frühen Entwicklungsphase. Die Klassifikation und Einschätzung der Pflegewissenschaft im Kontext ihrer wissenschaftssoziologischen, disziplinaren Bezüge ist laut Remmers (2011) dabei dependent von ihrer „empirischen Forschungsleistung“ (S. 10).

Statistische Methodik, als Teil dieser empirischen Performanz, unterliegt jedoch ebenso einer intra- und interdisziplinären Diskussion wie die Pflegewissenschaft selbst. In einer ersten Annäherung an eine disziplinbezogenen statistische Methodologie ist im Zuge einer Orientierung an bereits etablierten sozial- und naturwissenschaftlichen Verfahren eine bestimmte Tendenz zu einer frequentistischen Perspektive erkennbar (vgl. bspw. Müller, 2011). Die vorliegende Arbeit soll die methodische Perspektive der Pflegewissenschaft um eine bayesianische Statistik erweitern, über die oftmals als „klassisch“ (vgl. Tschirk, 2014) oder frequentistisch Bezeichnete hinaus. Interessanter Weise bietet der Ansatz der bayesianischen Statistik eine äußerst kompatible methodische Form für bestimmte inhaltliche Dialoge in der Pflegewissenschaft, wie beispielsweise das professionelle Handeln Pflegender. Eine besondere Relevanz entfaltet diese Arbeit aus Sicht des Autors jedoch vor dem Hintergrund einer stetig fortschreitenden epistemischen Diskussion in der Pflegewissenschaft und -forschung, die im Kontext einer philosophisch-methodologischen Selbstfindung zu interpretieren ist. Da, wie im Folgenden nachgewiesen wird, bisher keine übersichtliche, einführende Literatur für bayesianische Statistik im Kontext empirischer Pflegeforschung in Deutschland zu finden ist, erhebt diese Arbeit den Anspruch, grundlegend mathematische und verfahrensbezogene Statistik darzustellen. Jene kann also als eine Einführung in die bayesianische Statistik für die empirische Pflegeforschung verstanden werden. Dabei ist anzumerken, dass mathematische Kenntnisse, insbesondere in den Gebieten Algebra, sowie Integral- und Differentialrechnung einen großen Vorteil bei dem Verständnis detaillierter Operationalität unterschiedlicher statistischer

Verfahren verschaffen. Daher wird in dieser Arbeit auf eine Ausformulierung mathematischer Substanz in verständlicher Form Wert gelegt. Sollte sich die Disziplin der Pflegewissenschaft langfristig in modernen und angemessenen Methoden der empirischen Forschung beheimaten wollen, kann sie dies allerdings nicht unter Ausgrenzung logisch-formaler Expression tun.

Die vorliegende Arbeit gliedert sich wie folgt: Nach der Darstellung des methodischen Vorgehens und der systematischen Literaturrecherche soll ein kurzer historischer Abriss einer Entwicklung der bayesianischen Statistik erfolgen. Im Anschluss legt der Verfasser die ideelle und mathematische Grundlage eben jener bayesianischen Statistik dar. Wie gezeigt werden wird, sind diese beiden Prinzipien nicht voneinander zu trennen. Darauf aufbauend soll die mathematische und methodische Operationalisierung bestimmter statistischer Verfahren behandelt werden und ein Ergebnisvergleich angestellt werden zwischen einem frequentistisch und einem bayesianisch konstruierten Hypothesentest. Bevor dann eine Zusammenfassung vorgenommen wird, geht der Autor auf die bereits erwähnte intradisziplinäre Relevanz der Bayes-Statistik für die Pflegewissenschaft ein.

Der Autor möchte anmerken, dass im Folgenden der Einfachheit halber das generische Maskulinum verwendet wird. An den entsprechenden Stellen sind alle weiteren Geschlechter miteingeschlossen außer dieses ist gesondert im Text erwähnt.

2.Literaturrecherche und Methodik

Zur theoretischen Abbildung bayesianischer Statistik wurde eine systematische Literaturrecherche in den für die Pflegewissenschaft üblichen Datenbanken durchgeführt. Diese sind namentlich Gerolit, Livivo, Pubmed, die Cochrane Library und Cinahl. Für diese Datenbanken wurde ein Rechercheprotokoll angelegt. Es befindet sich im Anhang. Auf Grund der geringen Treffer wurden keine Ausschlusskriterien festgelegt. Ein im Protokoll erfolgter Ausschluss wurde immer auf Grund fehlenden inhaltlichen Bezuges und lediglich einmal wegen sprachlicher Barrieren durchgeführt. Neben den erwähnten Datenbanken wurde eine Suche im OPAC der Philosophisch-Theologischen Hochschule Vallendar und eine Handsuche nach dem Schneeballprinzip durchgeführt. Die Suche im OPAC ergab eine vernachlässigbare Anzahl an Treffern. Die Handsuche zeigte ebenso wie die systematische Recherche einen kleinen Kreis angemessener wissenschaftlicher Beiträge. Gezielte Suchen zu teils spezifischen Aspekten mathematischer Verfahren (wie bspw. eine genaue Definition einer bestimmten Wahrscheinlichkeitsverteilung) rundeten die Literaturlarbeit ab.

Im weiteren Verlauf dieser Arbeit wird eine Analyse bereits erhobener Daten durchgeführt. Sie dient lediglich zum besseren Verständnis der davor dargestellten Inhalte. Auch eine Analyse frei erdachter Daten wäre zu diesem Zweck geeignet, jedoch birgt die Bewertung realer Daten den Vorteil, Ergebnisse nicht durch unbewusste Beeinflussung leichter interpretierbar zu machen. Es wurden zwei Verfahren auf die Daten einer pflegerischen Interventionsstudie angewendet: Die Berechnung eines sog. Bayes-Faktors und eine Alternative zum klassisch frequentistischen T-Test, dem sog. BEST-Verfahren nach Kruschke (2013). Die Logik und genaue Vorgehensweise dieser Verfahren wird vor der exemplarischen Anwendung an entsprechender Stelle erklärt.

3.Forschungsstand

Die Sichtung von Monographien und Sammelwerken zeigte entweder eine sehr allgemeine und möglichst umfängliche Erklärung bayesianischer Verfahren mit einer Tendenz zu Unspezifiziertheit (z.B. Kaplan, 2014; Tschirk, 2014; Ziegler, 2017) oder sehr spezifischen Ansätzen bestimmter Aspekte bayesianischer Statistik, häufig mit einem Fokus auf mathematischer Herleitung und Differenziertheit (z.B. Phadia, 2016; Lecoutre & Poitevineau, 2014). Zwei Monographien wurden aus Sicht des Autors dem Anspruch einer allgemeinen, einführenden und gleichzeitig mathematisch gut interpretierbaren Ausführung gerecht (Cowles, 2013; Hoff, 2009). Eine etwas umfänglichere Darstellung der historischen Umstände und Entwicklung bayesianischer Statistik liefert Ziegler (2017).

Insgesamt 21 relevante anwendungsbezogene Artikel konnten im Rahmen der Literaturrecherche gefunden werden. 14 davon hatten als einen Schwerpunkt des Veröffentlichten eine methodische Erklärung der bayesianischen Statistik an sich. Diese blieben jedoch bis auf zwei Ausnahmen hinter der methodischen Differenziertheit der explorierten Monographien und Sammelwerken zurück. 13 der 21 Artikel hatten eine Ausrichtung auf klinische Phänomene, lediglich zwei Artikel waren auf das Setting der Altenpflege gerichtet. Zu ambulanten Pflegephänomenen wurde kein Artikel gefunden. Insgesamt bezogen sich jedoch nur neun der 21 Artikel auf tatsächliche pflegerische Phänomene (die übrigen unterlagen eher medizinischer und/oder methodischer Ausrichtung), während drei der neun pflegebezogenen Artikel auch die Thematik der Pflegequalität und somit tatsächlicher inhaltlicher Fragestellungen nachgingen, während die übrigen sechs Artikel ihren Fokus auf strukturelle oder organisationaler Erkenntnis legten. Auf die originäre Thematik bezogen können die folgenden Artikel tabellarisch eingeordnet werden:

Autor	Jahr	Schwerpunkt
Thompson & Martin	2017	Klinisch-methodisch
Peltikoski et al.	2015	Klinisch-strukturell
Tervo-Heikkinen, Kiviniemi, Partanen & Vehviläinen-Julkunen	2009	Klinisch-strukturell, organisational
Pitkäaho, Partanen, Miettinen & Vehviläinen-Julkunen	2015	Klinisch, organisational
Goodson, Jang & Rantz	2008	Altenpflege, Pflegequalität
Berlowitz et al.	2002	Altenpflege, strukturell, Pflegequalität
Lee & Abbott	2003	Methodisch
Lopes, da Silva & Araujo	2012	Methodisch
Sawka	2007	Pflegequalität, methodisch

Tab. 1: *Zusammenfassung aller pflegerelevanten Artikel mit Verwendung bayesianischer Verfahren*

Die in Tabelle 1 dargestellten Artikel können am ehesten vor dem Hintergrund eines methodischen Defizits in der Pflegewissenschaft kategorisiert werden. Selbst, bei den Artikeln, die ihren Schwerpunkt nicht auf das methodische Vorgehen legen, ist stets eine methodische Einführung mit der Vermittlung eher rudimentärer Kenntnisse bayesianischer Statistik zu finden. Tabelle 1 demonstriert darüber hinaus gut, wie selten eine Veröffentlichung von Artikeln, die bayesianische Statistik nutzen, überhaupt erfolgt. Insgesamt ergibt sich somit das Bild einer nicht weit verbreiteten Methodik, die insbesondere in Deutschland im Feld der Pflegewissenschaft anscheinend nicht sonderlich bekannt ist. Wie zum Ende dieser Arbeit gezeigt wird, hält diese Statistik jedoch ein gewisses Potential für die empirische Pflegeforschung bereit. Um die Hintergründe und die Operationalisierung bayesianischer Modelle zu durchdringen, soll vor der Erläuterung mathematischer Grundlagen zuerst ein historischer Abriss erfolgen. Die durch die Literaturrecherche gestützte

Annahme einer eher selten angewandten Statistik liegt unter anderem an dem Umgang mit ihr in der Vergangenheit.

4.Historische Betrachtung: Die Bayesianische Idee

Der Begriff „Statistik“ wird häufig im Kontext der Sozialwissenschaften als alleinstehende Entität zum Zweck einer empirischen Episteme definiert. Bortz & Schuster (2016) erklären in der Einleitung ihres umfänglichen Werkes „Statistik für Human- und Sozialwissenschaftler“:

„...die Statistik [stellt] den empirischen Wissenschaften Verfahren zur Verfügung, mit denen objektive Entscheidungen über die Brauchbarkeit von wissenschaftlichen Hypothesen getroffen werden können.“ (S.3).

Dass die Entwicklung der Statistik und ihrer heutigen Anwendung jedoch einer reflektierbaren Historie unterliegt und diskutable philosophische Überlegungen ihr eigenes Fundament bilden, ist meist nicht Thema eines solchen ‚Standardwerkes‘ (vgl. ebd.). Viel eher werden Wahrscheinlichkeitstheoreme unterschiedlicher Philosophien unter dem Anschein einer uniformen Theorie in einem Kapitel zusammengefasst (vgl. ebd. S. 49 ff.). Ausdruck eines philosophischen Fundaments gewinnt eine bestimmte Statistik insbesondere durch die Offenlegung ihres Wahrscheinlichkeitsbegriffs. Für Ziegler (2017) ist dieser der zentrale Ansatzpunkt einer Explikation bestimmter Grundannahmen einer jeweiligen statistischen Schule (vgl. S. 15 ff.).

Daher soll in diesem Kapitel die historische Entwicklung des Wahrscheinlichkeitsbegriffs nach Bayes und der frequentistische Wahrscheinlichkeitsbegriff (diese begriffliche Differenzierung übernimmt der Autor insb. von Lecoutre & Poitevineau (2014)) gegenübergestellt werden und so ein tieferes Verständnis für die keinesfalls feststehende und indiskutable Grundlage der modernen Statistik generiert werden.

4.1 Entwicklung vom 18. bis zum 19. Jahrhundert

Die Entwicklung und Diskussion einer Definition von Wahrscheinlichkeit im 18. Jahrhundert stellt sich nicht zwangsläufig sprachlich und begrifflich so manifest heraus, wie man annehmen mag. Viel mehr fußt sie nach Ziegler (2017) auf einer prinzipiellen, generellen Auseinandersetzung unterschiedlicher Intellektueller mit den Grundannahmen rationalen Handelns (was vor dem Hintergrund einer Zuwendung zu weltlichen Phänomenen im Zuge der Aufklärung interpretierbar ist) und einer dazu parallel laufenden Progressivität der mathematischen Operationalisierung von alltäglichen Fragestellungen (vgl. S. 24 ff.). Interessant ist hierbei, dass Rationalität im Kontext eines subjektiven Maßes an Vertrauen betrachtet wurde: Die Negation vollständiger Erklärungssicherheit über ein reales Phänomen und die Transformation der Vorstellung einer Objektivität in das Konstrukt der Mathematik (vgl. ebd.). Die dadurch anklingende Segregation von Subjektivität und Objektivität schien allerdings bei der Anwendung auf eine reale Situation nicht Teil des damaligen Denkens (zumindest in den Kreisen mancher Intellektueller) gewesen zu sein. Vielmehr wurde eine synergetische Erfahrung der Welt durch einen Grad an subjektivem Glauben einerseits und der Lumineszenz einer objektiven Existenz (als weltliche Manifestation) andererseits gedacht (vgl. ebd. S. 33 ff.). Der auch heutzutage teilweise direkt hergestellte Bezug von Rationalität und einer mathematisch-deterministischen Form ist eher als eine allmähliche Entwicklung zu betrachten (vgl. Gigerenzer, 1989, S.6 ff.). Als Beispiel kann die von David Hume (1711-1776) aufgestellte Theorie angeführt werden, dass die Wahrnehmung einer (objektiven) Welt von einem mit (subjektiver) Unsicherheit belasteten Ursache-Wirkungs-Prinzip abhängig ist und deshalb lediglich Aussagen über wahrscheinliche Zusammenhänge legitim seien (vgl. Ziegler, 2017, S. 33; vgl. McGrayne, 2012, S.5 ff.). Die entgegengesetzte Frage jenes Prinzips, also das Beobachten einer Wirkung und den Rückschluss auf die Ursache, war Anlass für Thomas Bayes (1702-1761), den Begriff der ‚bedingten Wahrscheinlichkeit‘ zu formulieren (ebd.). Im Vorfeld hatte sich Jakob Bernoulli (1654-1705) einem Wahrscheinlichkeitsbegriff definitorisch angenähert und formulierte darin Wahrscheinlichkeit als „Grad der Sicherheit“ (Bernoulli, 1713, S. 211 zitiert

nach Ziegler, 2017, S. 42). Auf diesem Begriff aufbauend legt Bayes die Überlegung von konditionellen Ereignissen dar:

„Er [Bayes] sucht nach der Wahrscheinlichkeit, die dem Eintreten eines konkreten empirischen Ereignisses zu zugrunde liegt, um sie dann auf ein zukünftiges Ereignis anwenden zu können.“ (Ziegler, ebd., S. 44).

Unabhängig von diesem ersten Fundament der heutigen bayesianischen Statistik entwickelte Pierre Simon Laplace (1749-1827) in den Grundzügen ein mathematisch operationalisiertes Äquivalent dieser Idee, bevor er mit Hilfe der niedergeschriebenen Gedanken Bayes' (die im Übrigen Richard Price, ein Freund Thomas Bayes' nach dessen Tod aufbereitete und veröffentlichte) Hypothesentests unternahm, die prinzipiell heutigen bayesianischen Operationen im Sinne einer Kombination von Vorwissen und neu explorierten Daten gleicht (vgl. McGrayne, 2012, S. 13 ff.). Anfang des 19. Jahrhunderts entwickelte Laplace eine mathematische Formel, die der in heutigen Lehrbüchern veröffentlichten Gleichung für bedingte Wahrscheinlichkeit oder dem *Satz von Bayes* (vgl. bspw. Büchter & Henn, 2007, S. 222) sehr ähnlich ist (vgl. Ziegler, 2017, S. 49; McGrayne, 2012, S.31).

4.2 Entwicklung ab dem 19. Jahrhundert

Ziegler (2017) beschreibt zu Anfang des 19. Jahrhunderts eine scharfe Kritik des Begriffs einer subjektiven Wahrscheinlichkeit oder eines Vertrauensgrades, die mit der Rezeption eines objektiven Wahrscheinlichkeitsbegriffs einhergeht (vgl. S. 53 ff.). In der Folge naturalistischer und deterministischer Überzeugung trete die ursprünglich nicht als Gegensatz gedachte Kombination eigenen Wissens und empirischer Beobachtung in den Hintergrund und das Bild einer objektiv erfahrbaren Welt unter der Voraussetzung mathematischer Beschreibbarkeit in den Vordergrund. Der subjektive Gehalt des Wahrscheinlichkeitsbegriffs und die damit einhergehende Aufforderung zur Darlegung der eigenen (theoretischen) Überzeugung wurden als

unwissenschaftlich oder irrational betrachtet, während die Fokussierung auf reine Daten (oder Zahlen) und deren Operationalität im Kontext aufkommender Sozialstatistiken einem objektiven Wahrheitsbegriff wesentlich näher kamen als eine bayesianisch motivierte Auslegung rationaler Wahrnehmung (vgl. ebd.). Die auf die Abbildung einer Norm gerichtete (Sozial)-Statistik des 19. Jahrhunderts stellt sich somit eher als beschreibende Disziplin dar:

„Methodisch und mathematisch geprägt wurde sie [die Zeit des 19. Jahrhunderts] durch eine reine Sichtweise der...deskriptiven Statistik. Als Analyseobjekt tauschte man den vernünftigen Menschen gegen den...Durchschnittsmenschen ein“ (Ziegler, 2017, S. 58).

In der zu dieser Zeit angewandten Statistik manifestierte sich diese im Kontext von Industrialisierung, Mechanisierung und Normalisierung in mathematischen Anwendungen wie Maße der zentralen Tendenz, Linearität und Normalverteilungen. Letztere, für die Laplace und De Moivre mit dem zentralen Grenzwertsatz ein Fundament schafften, wurde durch Gauß für die Darstellung von Wahrscheinlichkeiten bei ausreichend großen Stichproben legitimiert (vgl. Ziegler, 2017, S. 63 ff.). Dadurch verlor allerdings eine theoretische Implikation oder Vorannahme für die statistische Analyse eines Sachverhalts an Bedeutung und ermöglichte eine Separierung von theoretischer Betrachtung und empirischer Analyse (vgl. ebd.). Die Vorstellung einer unfehlbaren objektiven Statistik (vgl. Gigerenzer, 1989, S. 39) supprimierte in Folge dessen allerdings eine wissenschaftlich-methodische Diskussion. Parallelen zu heute lassen sich bspw. im Fachgebiet der Medizin aufzeigen: Die statistische Analyse klinischer Versuche kann an eine nur für die Auswertung zuständige Institution, die Biometrie, weitergegeben werden. Inhaltliche Nachvollziehbarkeit oder Relevanz stellt dabei jedoch keine notwendige Bedingung für die methodische Auswertung der Biometrie dar. Die Anfechtbarkeit bestimmter mathematisch-statistischer Methoden oder Modellen per se gestaltet sich schwierig, da Ungenauigkeiten oder unbefriedigende Ergebnisse nicht direkt auf jene bezogen werden.

Über die Analyse von Varianz und Abweichungen vom Mittelwert kristallisierte sich zu Anfang des 20. Jahrhunderts die Grundlage einer

heutigen Inferenzstatistik. Francis Galton und Karl Pearson fragten im Kontext der Sozialstatistiken nach Ursachen, Zusammenhängen und Erklärungsmodellen auftretender Variation. Gerade Karl Pearson umging dabei jedoch die für Thomas Bayes noch überaus wichtige Diskussion von Ursache-Wirkungs-Zusammenhängen (bzw. deren Inversion) und konstruierte eine objektiv anmutende Darstellung eines Zusammenhangs in Form von Korrelationen (vgl. ebd., S. 72 ff.). Dabei wandelte sich in der Statistik die Perspektive auf Phänomene der realen Welt von einer Deterministischen zu einer Probabilistischen: Die Verwendung von Residuen sprach statistischen Modellen (bspw. die der linearen Regression) eine gewisse Unschärfe zu. Die Interpretation dieser Unschärfe wurde jedoch nicht vor dem Hintergrund einer subjektiven Unsicherheit des Forschers oder dessen theoretisches Konstrukt diskutiert, sondern als ein partielles Relikt der realen Welt betrachtet, die im weiteren Forschungsprozess Bestimmtheit erfahren kann. Eine solche Betrachtung trat somit zur scheinbaren Objektivierung der Disziplin der Statistik bei (vgl. ebd.). Auf dieser Grundlage entstand eine wissenschaftliche Diskussion um die Methodik des Schließens von einer Stichprobe auf die Grundgesamtheit, im Wesentlichen ausgetragen durch Egon Pearson und Jerzy Neyman einerseits und Ronald Aylmer Fisher andererseits.

Fisher argumentierte für das Testen einer Nullhypothese (H_0) auf Grundlage empirischer Daten. Bei einer ausreichend kleinen Wahrscheinlichkeit (heutzutage als p -Wert oder Signifikanzniveau beschrieben) der Nullhypothese sollte diese verworfen werden (die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, obwohl sie richtig ist, bezeichnet man als α -Fehler). Aus diesem besonders in t -Tests oder z -Tests mit Hilfe des Vergleichs von Mittelwerten abbildbaren Procedere resultiert heute eine weit verbreitete, teilweise unreflektierte Akzeptanz bestimmter Signifikanzniveaus (5% oder 1%; Fisher war tatsächlich der Meinung, dass festgelegte Signifikanzniveaus nicht sinnvoll waren) und dem Testen teils atheoretischer Nullhypothesen (bspw. die in der Medizin häufig vorkommende Annahme, dass eine bestimmte Intervention keinen Effekt hat) (vgl. ebd., S. 90 ff.). Neyman und Pearson thematisierten auf diesen Vorschlag Fishers das damit verbundene Problem, durch das Verwerfen einer Nullhypothese trotzdem keine positive inhaltliche Aussage

über einen Gegenstand treffen zu können. Als Lösung boten Neyman und Pearson an, für den Test empirischer Daten mindestens zwei Hypothesen aufzustellen: Die bereits von Fisher formulierte H_0 -Hypothese, sowie eine inhaltliche Hypothese, heute üblicherweise als H_1 -Hypothese bezeichnet. Neben dem α -Fehler führten Neyman und Pearson konsequenter Weise einen zweiten Fehler oder eine zweite Fehlerwahrscheinlichkeit ein: Die Wahrscheinlichkeit dafür, die Nullhypothese beizubehalten, obwohl sie falsch ist (der sog. β -Fehler).

Der wesentliche Unterschied beider Schulen besteht folglich darin, dass Fisher das Ablehnen bestimmter Annahmen als Schritt auf einem Erkenntnis-Kontinuum begriff (im Sinne eines kumulativen Lernprozesses), während Neyman und Pearson das Testen bestimmter Annahmen (im besten Fall aller Annahmen zu einem bestimmten Phänomen) als abgeschlossenes, mathematisch inspiriertes Konstrukt betrachteten, nach dessen Bearbeitung eine eindeutige Aussage über den Gegenstand zu treffen sei (vgl. ebd.). Ab der zweiten Hälfte des 20. Jahrhunderts wurden beide statistischen Schulen im Kontext des Einzugs der Statistik in die universitäre Lehre pragmatisch miteinander kombiniert und führten so zu einer dogmatisch anmutenden Vereinheitlichung frequentistischer Inferenz. Theoretische Annahmen über Wahrscheinlichkeit, die trotz dem Versuch der Objektivierung ab dem 19. Jahrhundert bis zur ersten Hälfte des 20. Jahrhunderts expliziert und diskutiert wurden, traten damit in den Hintergrund (vgl. ebd.). Der Wiedereintritt der bayesianischen Statistik wurde durch die Rezeption dieser statistischen Schule durch wenige Personen in unterschiedlichen wissenschaftlichen Disziplinen (z.B. Frank Ramsey [Philosoph] oder Bruno de Finetti [Mathematiker]) und der mathematisch-operationalen Umsetzbarkeit durch die Entwicklung der Informationstechnologie ab den 80'er Jahren des 20. Jahrhunderts ermöglicht (vgl. ebd., S. 110 f.).

Zusammenfassen lässt sich die historische Entwicklung der bayesianischen und frequentistischen Statistik wie folgt: Die dargestellte Perspektive einer Synergie von Subjektivität und Objektivität, die im 18. Jahrhundert und davor gedacht wurde, wurde durch die künstliche Vermischung beider erkenntnistheoretischer Ansätze im Laufe des 19. und

die Ablehnung einer subjektiven Statistik im ersten Teil des 20. Jahrhunderts aufgehoben. Die Folge ist eine für die frequentistische Statistik charakteristische Methode möglichst atheoretischer und ‚unbefangener‘ Haltung des Forschers. Dadurch wird die Illusion geschaffen, tatsächlich objektive Erkenntnisse aus einem empirischen Sachverhalt ableiten zu können. Gerade bei häufig verwendeten Methoden wie bspw. dem T-Test, der ANOVA (kurz für: Analysis of variance) oder (generalisierten) linearen Regressionen ist die Formulierung bereits vorhandenen Wissens nicht oder nur selten (z.B. bei der a priori Fallzahlschätzungen mit Hilfe einer angenommenen Effektstärke auf Grundlage einer T-Statistik) Teil des Prozederes, geschweige denn wichtiger Faktor für die Interpretation des Ergebnisses. Eine fehlende Diskussion innerhalb oder außerhalb der frequentistischen Statistik ist derweil zurück zu führen auf die pragmatische Mechanisierung möglichst handlungsnaher empirischer Analysen in der zweiten Hälfte des 20. Jahrhunderts. Das momentan zu beobachtende Aufkommen der bayesianischen Statistik in unterschiedlichen Handlungsbereichen ist in erster Linie der Arbeit Weniger und dem informationstechnischen Fortschritt geschuldet.

5. Mathematische Grundlagen und Operationalisierung

In diesem Kapitel sollen Grundlagen und Verfahren der bayesianischen Statistik, insbesondere der bayesianischen Inferenzstatistik dargelegt werden. Dies ist aus Sicht des Autors einerseits notwendig für die Nachvollziehbarkeit mathematisch automatisierter Verfahren (verschiedener Statistik-Programme, wie bspw. R), andererseits soll im weiteren Verlauf der Arbeit an Hand eines empirischen Beispiels die Differenz eines klassisch-frequentistischen und eines bayesianischen Verfahrens aufgezeigt werden. Dieser Abschnitt der Arbeit unterteilt sich in einen ersten, allgemeinen Teil und dient insbesondere einem fundamentalen Verständnis bayesianischer Ansätze. Im zweiten Teil werden konkrete bayesianische Verfahren erläutert. Dazu zählen Bayes-

Faktor, bayesianisches Hypothesentesten, generalisierte lineare Modelle und das bayesianische Äquivalent der Mehrebenenanalyse.

5.1 Grundsatz

Der *Satz von Bayes* oder zutreffender der *Satz der bedingten Wahrscheinlichkeiten* (da ihn wie bereits erwähnt Bayes in persona weder in einer mathematisch eindeutigen Form niedergeschrieben noch veröffentlicht hatte) entspricht in seiner heutigen prinzipiell der von Laplace entwickelten Form:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Dabei ist y eine beliebige bekannte Variable, θ (Theta) ein beliebiger unbekannter Parameter, $p(\theta|y)$ die sog. *Posteriori-Verteilung*, $p(y|\theta)$ die Wahrscheinlichkeit einer Variablen y gegeben eines Parameters θ , die sogenannte *Likelihood*, des Weiteren $p(\theta)$ die sog. *Priori-Verteilung*, und $p(y)$ ein normalisierender Term (vgl. Kaplan, 2014, S. 13 ff.).

Bevor die einzelnen Terme genauer beschrieben werden, soll an dieser Stelle auf die epistemische Äquivalenz dieses Satzes mit der Idee einer synthetischen subjektiv-objektiven Rationalität hingewiesen werden. Denn letztendlich besagt der *Satz der bedingten Wahrscheinlichkeiten*, dass sich ein Wissen über ein Phänomen (*Posteriori*) aus dem ergibt, was der Forscher bereits vorher darüber wusste (subjektives *Priori*) und dem, was er an Hand seiner Untersuchung(en) darüber herausgefunden hat (empirisch kalkulierte Wahrscheinlichkeit einer Variablen y gegeben eines Parameters θ) im Kontext dessen, was als möglicher Wissensraum gedacht werden kann (normalisierender Term). Die Offenlegung eines persönlichen Wissens (was wohl als Überzeugungs- oder Glaubensgrad

definiert werden kann) ist also essentieller Bestandteil der bayesianischen Statistik.

5.1.1 *Priori-Verteilung*

Die *Priori*- oder *a priori*-Verteilung beschreibt im Sinne der bayesianischen Idee das Wissen oder den Glauben über das Wissen hinsichtlich eines bestimmten Phänomens und dessen Zusammenhang. Kaplan (2014) unterteilt zwei Arten einer *Priori*-Verteilung (engl. *priors*): Informative und nicht-informative *priors* (vgl. S. 18 ff.). Bei nicht-informativen *priors* handelt es sich um Verteilungen, die keinen Informationsgehalt aus früheren phänomenspezifischen Erkenntnissen enthalten. In diesem Zusammenhang werden insbesondere Gleichverteilungen erwähnt (vgl. ebd., S. 19; vgl. Bättig, 2017, S. 316), die mitunter auch in Programmen in *R* als default-setting benutzt werden (vgl. Kaplan, 2014, S.19). Die Gleichverteilung eignet sich zum Ausdruck der Vermutung, dass Werte eines bestimmten Bereichs $[a,b]$ mit der gleichen Wahrscheinlichkeit auftreten:

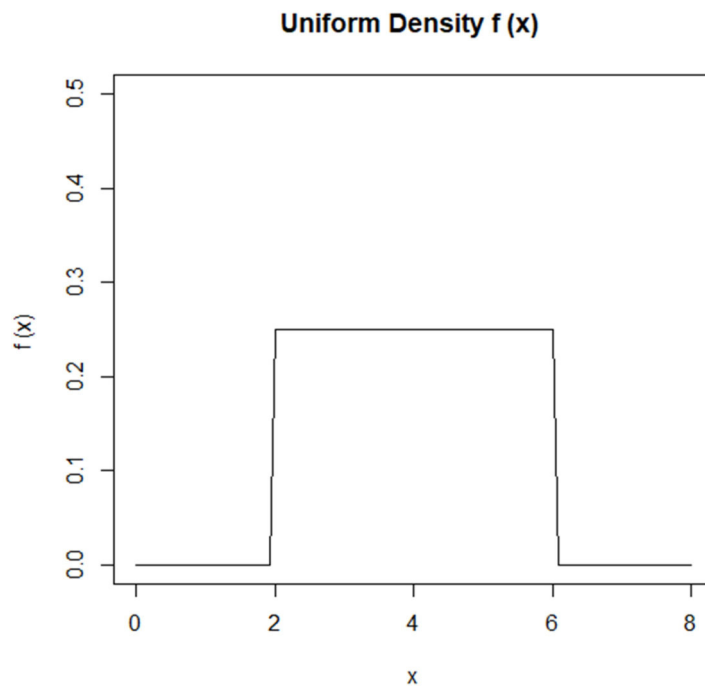


Abb.1: Graph einer stetigen Gleichverteilung mit $f(x=0,25)$ im Wertebereich $[2,6]$.

Aus Abbildung 1 ist erkennbar, dass sich das Vorwissen über den hypothetischen Sachbereich darin erschöpft, dass stetige Werte im Bereich $[2,6]$ jeweils mit einer Wahrscheinlichkeit $f(x) = p_x = 0,25$ auftauchen. Die integrierte Wahrscheinlichkeitsfunktion $\int_2^6 f(x) dx = [0,25x]_2^6$ ergibt 1 über dem für sie definierten Wertebereich. Die Integration von kontinuierlichen Variablen und die Summierung von diskreten Variablen zu 1 bezeichnet nach Kaplan eine notwendige Voraussetzung für jede beliebige Form der a-priori Wahrscheinlichkeitsverteilung (vgl. ebd.). Held & Bové (2014) bezeichnen nicht-informative *priors* im Grenzbereich als „improper prior distribution[s]“ (S. 184). Diese bezeichnen *Priori*-Verteilungen, die summiert, bzw. integriert $\neq 1$ ergeben. Diese ‚ungenauen‘ *priors* werden von Held & Bové (ebd.) als legitim definiert, falls eine daraus resultierende *Posteriori*-Verteilung trotzdem Anspruch genügt, dass das Integral oder die Summe ihres Definitionsbereichs 1 ergibt. Nicht-informative *priors* können darüber hinaus in konjugierte und nicht-konjugierte *priors* unterteilt

werden (vgl. Kaplan, 2014, S. 19 f.). Konjugierte *priors* besitzen die Eigenschaft, dass ihre Verteilungs- oder Funktionsart der der *Posteriori*-Verteilung entspricht. Aus der Problematik, dass die bereits erwähnten, häufig als nicht-informative *priors* genutzten Gleichverteilungen im Falle einer Verteilungstransformation zu teils stark variaten Funktionen werden (vgl. ebd.; Held & Bové, S. 185), entwickelte Harold Jeffreys eine Bedingung für nicht-informative, konjugierte *priors*, die in der Form

$$p(\theta) \propto \sqrt{[I(\theta)]}$$

als *Jeffreys' prior* bezeichnet wird, wobei $[I(\theta)]$ die sog. *Fisher Informationsmatrix*¹ bezeichnet (vgl. Kaplan, 2014, S. 20 f.). Dabei gelten diese Arten von *priors* nicht nur im Kontext von Gleichverteilungen. Folgende Tabelle gibt eine Übersicht über verschiedene Verteilungen, die in diesem Kapitel im Verlauf erläutert werden:

Likelihood-Funktion	Jeffreys' prior
Binomial	Beta
Geometrisch	Beta
Poisson	Gamma
Exponential	Gamma
Normal (Varianz bekannt)	Normal
Normal (Varianz unbekannt)	Inverse Gamma ²

Tab. 2: *Jeffreys' prior bei unterschiedlicher Likelihood, n. Held & Bové, 2014, S. 189.*

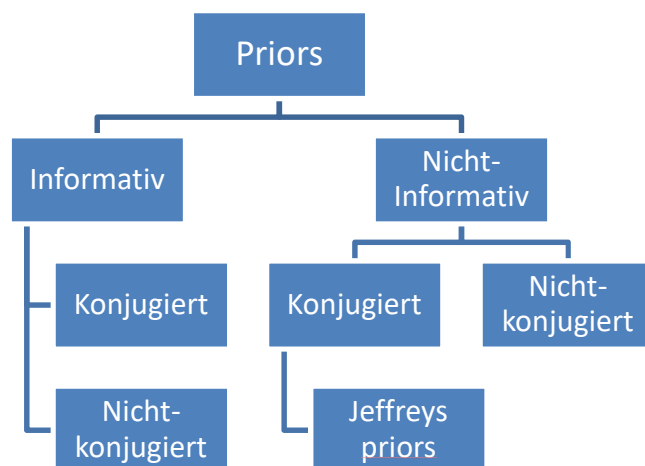
Informative *priors* enthalten hingegen wie zu vermuten ein bestimmtes Vorwissen über das zu untersuchende Phänomen. Kaplan (ebd.) unterteilt wiederum auch die informativen *priors* in konjugierte und nicht-konjugierte

¹ Letztendlich dient die Fisher-Informationsmatrix dem Auffinden eines anhand der Varianz gemessenen optimalen Schätzmodells für eine unbekannte Variable aus einer bekannten Variablen (vgl. Brokate et al., 2016, S. 913).

² Eine inverse Funktion f^{-1} oder auch Umkehrfunktion genannt, ordnet jedem ihrer Elemente ein ursprüngliches Element der Funktion f zu. Eine ausführliche Definition findet sich bspw. in Neher (2018), S. 139.

Wahrscheinlichkeitsverteilungen (ebd.). Die Konstruktion konjugierter *priors* aus Gründen der reinen Pragmatik und nicht aus Gründen der möglichst genauen Abbildung des Vorwissens kann sicherlich kritisch betrachtet werden. Informative *priors* berücksichtigen im Gegensatz zu nicht-informativen jedoch besonders die subjektivistische Natur der bayesianischen Statistik durch die Transparenz eigenen Wissens.

Das nachfolgende Diagramm soll eine zusammenfassende Übersicht verschiedener *prior*-Arten geben:



Diagr. 1: *Verschiedene priors und ihre Zusammenhänge (eigene Darstellung).*

Lecoutre (2014) verweist darauf, wie essentiell und kontrovers sich die Auswahl eines angemessenen *priors* darstellt, bspw. in der Frage nach legitimen Informationen, aus denen dieser gebildet wird (S. 13 f.). Solche Überlegungen münden in einer Diskussion über Wissensordnung und –hierarchisierung, die sich im pflegewissenschaftlichen Kontext simultan bspw. durch die Differenzierung qualitativer und quantitativer Forschung per se, der EBN/EBM-Debatte (*evidence based nursing / evidence based medicine*) oder Literaturbewertungssysteme (JBI appraisal tools, PRISMA, etc.) ausdrückt.

Um informative *priors* oder im Besonderen konjugierte *priors* abzubilden, ist es notwendig, die darin enthaltene Information mathematisch zu

modellieren und in eine funktionale Verteilung zu überführen. Die für die bayesianische Statistik wichtigsten Verteilungsformen (neben der bereits erwähnten Gleichverteilung) sollen in Form eines Einschubs nachfolgend kurz beschrieben werden.

5.1.2 Wahrscheinlichkeitsverteilungen

Das Modellieren unterschiedlicher Verteilungsfunktionen beinhaltet in der bayesianischen Statistik insbesondere vor dem Hintergrund der Operationalisierung von *priors* eine hohe Relevanz. Eine mathematisch reformulierte Transparenz des eigenen Vorwissens entspricht damit der Essenz eines subjektivistischen Wahrscheinlichkeitsbegriffes. Die folgenden Verteilungsfunktionen sind vor diesem Hintergrund zur Abbildung des eigenen, bzw. aus der Literatur exzerpierten Wissens notwendig, haben jedoch nicht zwangsläufig die ausschließliche Rolle einer *Priori*-Verteilung. Die aus den empirischen Daten abbildbare Likelihood oder die aus dieser und der *Priori*-Verteilung ableitbare *Posteriori*-Verteilung sollten ebenfalls einer mathematischen Genauigkeit genügen. In der vom Autor explorierten Literatur wird deutlich, dass insbesondere die Wahl bestimmter Verteilungen als *prior* von der Verteilung der *Likelihood* abhängig ist (vgl. Kaplan, 2014; Held & Bové, 2014). Es soll für die folgenden grafischen Darstellungen bestimmter Funktionen darauf hingewiesen werden, dass für eine einheitliche Abbildung alle Funktionen als stetige Verteilungen simuliert wurden, obgleich einige Verteilungen diskret sind.

Normalverteilung

Die allgemeine Form der stetigen Normalverteilung kann beschrieben werden durch

$$p(y|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2}$$

im Wertebereich $y \in [-\infty, \infty]$ (vgl. Hoff, 2009, S. 67) mit dem Erwartungswert θ und der Varianz σ^2 . Abbildung 2 entspricht einer grafischen Realisation.

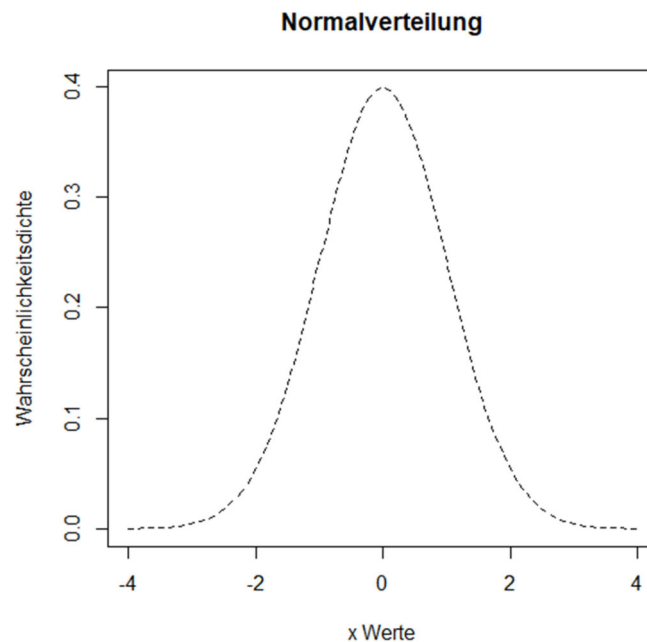


Abb. 2: Dichtefunktion einer Normalverteilung mit $\theta=0$, $\sigma^2=1$ im Bereich $[-4,4]$.

Die Form der Normalverteilung wurde bereits durch Karl Friedrich Gauß beschrieben. Im 19. Jahrhundert entdeckte Adolphe Quetelet die Approximation physiologischer Merkmale an die Normalverteilung (vgl. Schäfer, 2010, S. 93). In der klassisch-frequentistischen Statistik stellt die Normalverteilung eine wichtige Voraussetzung für bestimmte Parameter unterschiedlicher Verfahren dar, wie bspw. dem T-Test, der einfaktoriellen Varianzanalyse (ANOVA) oder der linearen Regression.

Bei bekanntem σ^2 und unbekanntem θ ergibt sich die *Priori*-Verteilung von θ als

$$p(\theta|\theta_0, \sigma_0^2) \propto \frac{1}{\sigma_0} e^{-\frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma_0}\right)^2}$$

(vgl. Kaplan, 2014, S. 35; Held & Bové, 2014, S. 180) mit θ_0, σ_0^2 als sog. „hyperparameters“ (Kaplan, 2014, S. 35), sprich den a priori angenommenen Werten für θ und σ^2 .

Bei unbekanntem σ^2 und bekanntem θ ergibt sich die *Priori*-Verteilung von σ^2 als

$$p(\sigma^2) = (\sigma^2)^{-(\alpha+1)} e^{-\frac{\beta}{\sigma^2}}$$

(vgl. ebd., S. 36) mit $\alpha, \beta > 0$, wobei α Gestaltparameter und β Längenparameter sind. Aus einem normalverteilten *prior* und einer normalverteilten *Likelihood* resultiert eine ebenfalls normalverteilte *Posteriori*-Verteilung, wie Held & Bové (2014) zeigen (vgl. S. 181).

Geometrische Verteilung

Die von Büchter & Henn (2007) als Lösung für „Wartezeitprobleme“ (S. 314) dargestellte geometrische Verteilung macht eine Aussage über die Anzahl der Wiederholungen eines Ereignisses, nach derer ein bestimmter Variablenwert (klassischer Weise der ‚Erfolg‘ eines Experiments) eintritt. Die geometrische Verteilung besitzt die Gleichung

$$P(X = k) = (1 - p)^{k-1} * p$$

(vgl. ebd.), wobei p die Wahrscheinlichkeit eines bestimmten Ergebnisses (‚Erfolg‘) ist und k die Anzahl der hierfür benötigten Wiederholungen. Abbildung 3 veranschaulicht den Verlauf eines typischen wiederholten Ereignisses in Abhängigkeit der jeweiligen Auftrittswahrscheinlichkeit.

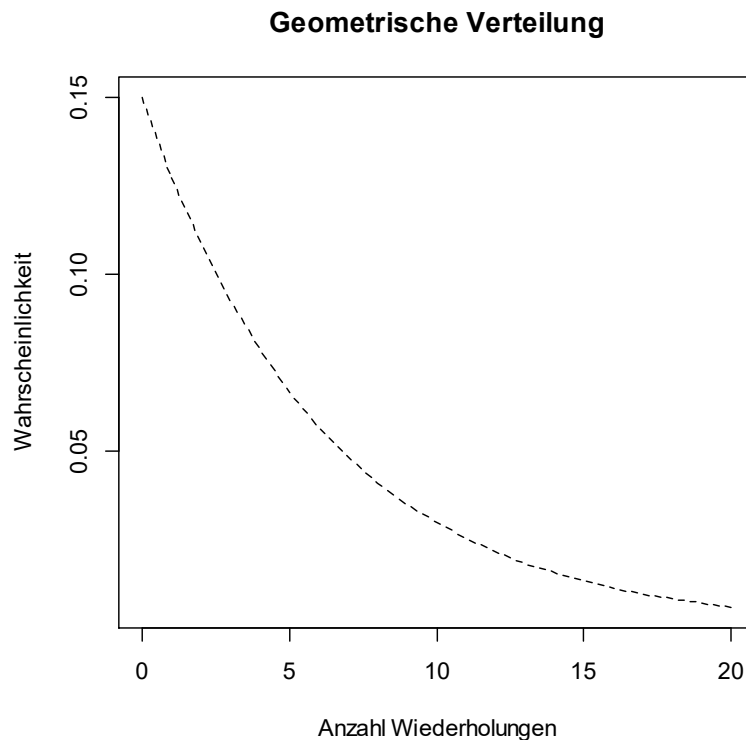


Abb. 3: Geometrische Verteilung mit $p = 0,15$ über dem Intervall $[0, 20]$.

Der konjugierte *prior* einer geometrisch verteilten Funktion ist die Beta-Verteilung, welche im Anschluss an die Binomialverteilung in diesem Kapitel erläutert wird.

Binomialverteilung

Die mathematische Abbildung eines sog. *Bernoulli*-Experiments, also die Folge von Ereignissen, die exakt zwei Merkmalsausprägungen annehmen können, wird Binomialverteilung genannt (vgl. Büchter & Henn, 2007, S. 306 f.). Die Wahrscheinlichkeitsdichtefunktion lautet

$$p(y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

, wobei n die Anzahl der gesamten Ereignisse (z.B. ‚Erfolg‘ und ‚Misserfolg‘), y die Anzahl der Ereignisse genau einer

Merkmalsausprägung (z.B. ‚Erfolg‘) und θ die Wahrscheinlichkeit des Ereignisses genau einer Merkmalsausprägung (z.B. ‚Erfolg‘) darstellt (vgl. Kaplan, 2014, S. 40 f.). Durch den Satz von *Moivre-Laplace* konnte gezeigt werden, dass die Binomialverteilung bei großem n der Normalverteilung entgegen strebt. Abbildung 4 zeigt eine Binomialverteilung von Erfolgen in Abhängigkeit der Ereigniswahrscheinlichkeit bei $n = 100$ und $\theta = 0,5$ im Bereich $y \in [20,80]$.

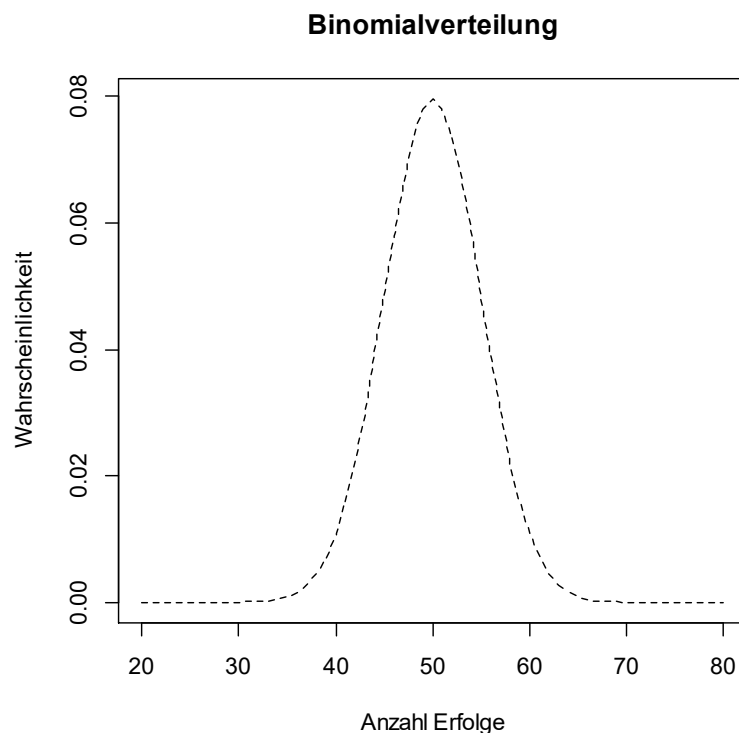


Abb.4: Binomialverteilung mit $\theta = 0,5$ im Intervall $[20,80]$.

Als konjugierter *prior* der Binomialverteilung dient die sog. Beta-Verteilung. Kaplan (ebd.) bezeichnet sie als „one of the most important distributions in statistics...“ (S. 41).

Beta-Verteilung

Die Beta-Verteilung ist eine aus Gamma-Verteilungen (s. dieses Kap.) kombinierte Funktion der Form

$$f(\theta|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

mit α (>0) und β (>0) als Längen- und Gestaltparameter (vgl. ebd., S. 41 f.). θ bezeichnet ebenso wie bei der Binomialverteilung die Wahrscheinlichkeit des Auftretens einer bestimmten Merkmalsausprägung („Erfolg“) durch das Verhältnis $\frac{y}{n}$, sprich die Anzahl Ereignisse einer bestimmten Merkmalsausprägung geteilt durch die Anzahl aller Ereignisse. Die Beta-Verteilung ist im Intervall $[0,1]$ definiert (vgl. ebd.). Beim Vorliegen der *Likelihood* als Binomialverteilung oder Geometrische Verteilung entspricht der konjugierte *prior* dem der Beta-Verteilung (vgl. Held & Bové, 2014, S. 181). Abbildung 5 visualisiert eine typische Beta-Verteilung:

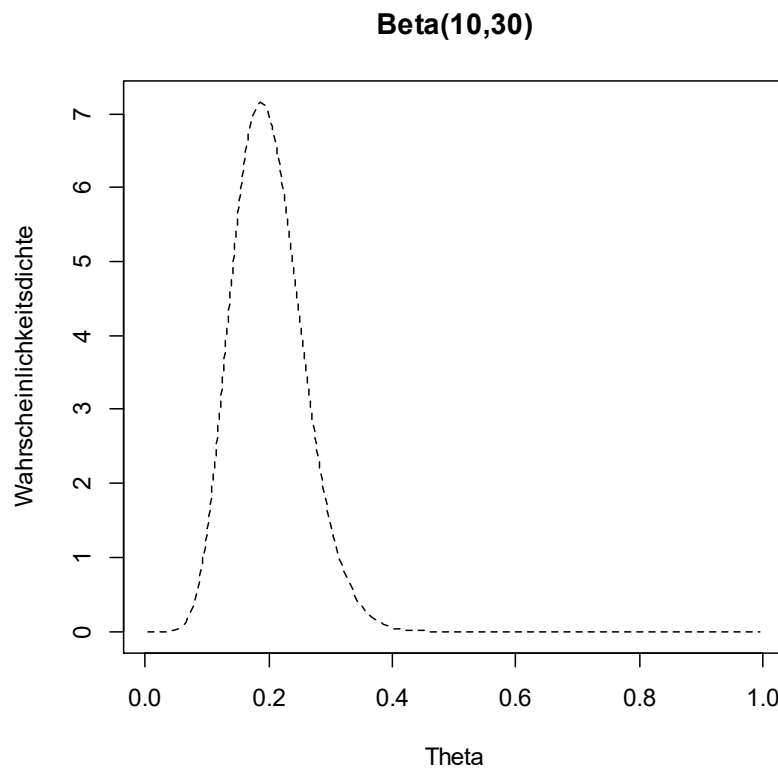


Abb.5: Beta-Verteilung mit $\alpha = 10$ und $\beta = 30$ im Bereich $\text{Theta} = \theta \in [0,1]$.

Kaplan beschreibt Beta-Verteilungen als äußerst flexibel: Je nach Wahl der Parameter (vgl. ebd.) α und β beeinflusst jene die *Posteriori*-Verteilung in ähnlicher Weise wie ein normalverteilter, ggf. aber auch wie ein gleichverteilter *prior*.

Multinomialverteilung

Die Multinomial-Verteilung entspricht einer generalisierten Form des bereits erwähnten Bernoulli-Experiments mit mehr als zwei möglichen Ereignissen (vgl. Büchter & Henn, 2007, S. 310). Die Dichtefunktion der Multinomialverteilung ergibt sich aus dem Multinomialkoeffizient

$$\frac{n!}{k_1! k_2! \dots k_s!}$$

, wobei n die Anzahl der Ausführungen eines bestimmten Experiments ist (bspw. die Anzahl eines beantworteten Items Y aller Fragebögen) und $k_1 k_2 \dots k_s$ die Anzahl bestimmter Ereignisse (bspw. die Anzahl des mit Wert x beantworteten Items Y), sowie der einzelnen Auftrittswahrscheinlichkeiten eines bestimmten Ereignisses $p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}$:

$$P(X_1 = k_1 \wedge X_2 = k_2 \wedge \dots \wedge X_s = k_s) = \frac{n!}{k_1! k_2! \dots k_s!} p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}$$

(vgl. Büchter & Henn, 2007, S. 310 f.). Die Multinomialverteilung dient insbesondere zur Kalkulation und Darstellung mehrdimensionaler Funktionen. Abbildung 6 zeigt eine Verteilung im Kontext von Marktrisikoschätzungen. Sie dient in diesem Kontext in erster Linie der bildlichen Veranschaulichung multinomialer Verteilungen.

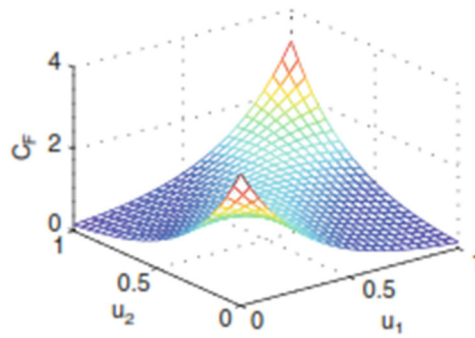


Abb. 6: Verteilung einer zweidimensionalen Copula (Misiorek & Weron, 2012, S. 1049).

Der konjugierte *prior* einer Multinomialverteilung ist die sog. Dirichlet-Verteilung (vgl. Kaplan, 2014, S.44). Diese gibt im Gegensatz zu multinomialen Verteilungen nicht die Wahrscheinlichkeitsverteilung eines wiederholten Ereignisses an, sondern Wahrscheinlichkeiten bestimmter Wahrscheinlichkeitsverteilungsfunktionen. Die Erläuterung und Darstellung der Dirichlet-Verteilung geht über den Rahmen dieser Arbeit hinaus. Eine ausführliche Darlegung dessen ist bspw. bei Phadia (2016) zu finden.

Poisson-Verteilung

Die nach Siméon Denis Poisson benannte Verteilung stellt eine Approximation an die Binomialverteilung dar. Jene bestimmt bei relativ großen Wiederholungen n eines Experiments und relativ kleiner Auftretswahrscheinlichkeit p eines bestimmten Ergebnisses wie oft eben jenes Ergebnis k -mal eintritt. Die Verteilungsfunktion lautet

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

mit $\lambda = n * p$ (vgl. Arens et al., 2015, S. 1453). Abbildung 7 zeigt eine Poisson-Verteilung von k in Abhängigkeit der Auftretswahrscheinlichkeit mit $\lambda = 10$.

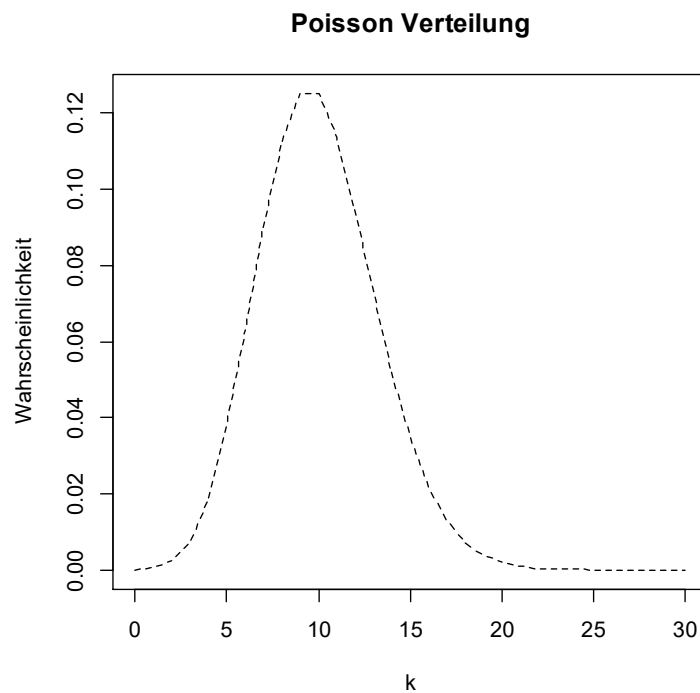


Abb.7: Poisson-Verteilung mit $\lambda = 10$ im Intervall $[0, 30]$.

Der konjugierte *prior* der Poisson-Verteilung ist die sog. Gamma-Verteilung. Diese ist wie bereits gezeigt eng mit der Beta-Verteilung verwandt.

Gamma-Verteilung

Die Gamma-Verteilung generalisiert Elemente der Fakultät $n!$, wobei $n \in \mathbb{N}$ ist, für den reellen und sogar für den komplexen Zahlenbereich (vgl. ebd., S. 1278). Für den Poisson-verteilten Parameter θ , für den gilt

$$P(X = k) = \frac{\theta^k}{k!} e^{-\theta}$$

ergibt sich die als konjugierter *prior* fungierende Gamma-Funktion

$$g(\theta) = \theta^{\alpha-1} e^{-\beta\theta}$$

(vgl. Kaplan, 2014, S. 40). α und β sind hierbei äquivalent zur Beta-Verteilung Gestalt- und Längenparameter. Eine allgemeinere Form beschreibt die Gleichung

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

(vgl. Arens et al., 2015, S.1278) für alle $x > 0$, $x \in \mathbb{R}$. Abbildung 8 zeigt eine Gamma-Verteilung mit Längenparameter $\alpha = \frac{1}{3}$ und Gestaltparameter $\beta = 4$:

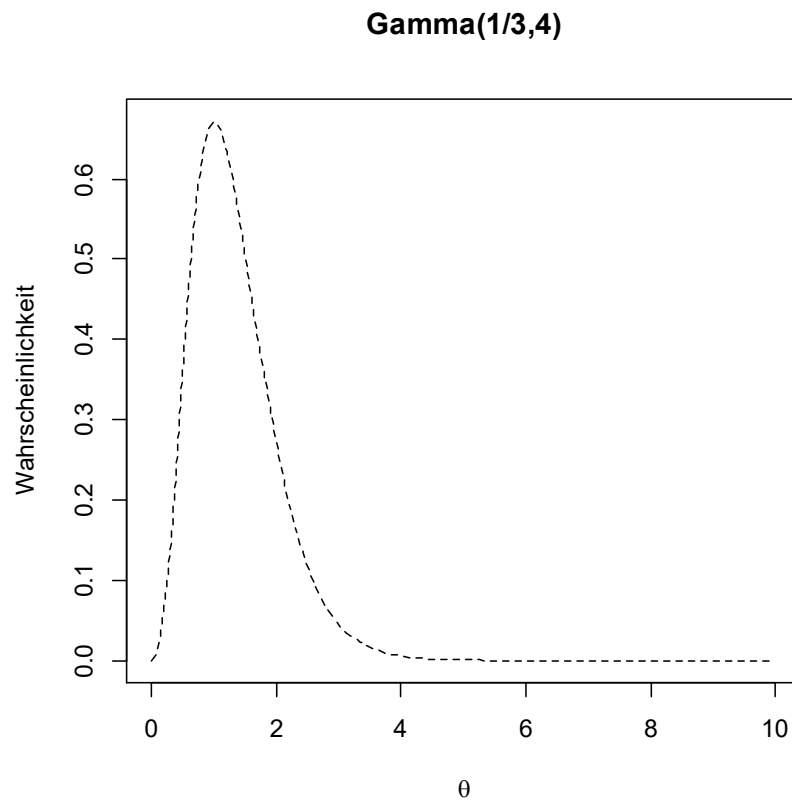


Abb. 8: Gamma-Verteilung für $\alpha = \frac{1}{3}$; $\beta = 4$ im Intervall $[0, 10]$.

Die Gamma-Funktion dient nicht nur für die Poisson-Verteilung als konjugierter *prior*, sondern auch für Exponentialfunktionen (vgl. Held & Bové, S. 181).

Cauchy-Verteilung

Da im folgenden Kapitel an Hand eines empirischen Beispiels potentielle Differenzen beim Einsatz klassisch frequentistischer und bayesianischer Analyse aufgezeigt werden sollen, wird zu guter Letzt die sog. Cauchy-Verteilung erläutert, die Teil der Berechnung des sog. „Bayes Factor“ (vgl. Morey & Rouder, 2015) mit Hilfe des Statistikprogramms R (R Core Team, 2016) ist.

Die Cauchy-Verteilung (nach dem Mathematiker Augustin Louis Cauchy) entstand ursprünglich aus dem mathematischen Problem, eine Dichtefunktion für den Quotient zweier Dichtefunktionen mit unabhängigen Variablen anzugeben (vgl. Brokate, 2016, S. 831). Die Cauchy-Verteilung besitzt die Form

$$p(x_i|x_0, y_0) = \frac{y_0}{\pi[y_0^2 + (x_0 - x_i)^2]}$$

(vgl. Hanson & Wolf, 1996, S. 255), wobei y_0 und x_0 Konstanten sind, die Streubreite und Lage der Funktion im Koordinatensystem beeinflussen und π die Kreiszahl pi darstellt. Abbildung 9 zeigt eine Cauchy-Verteilung mit dem Lageparameter $x_0 = 2$ und dem Streuungsparameter $y_0 = 1$.

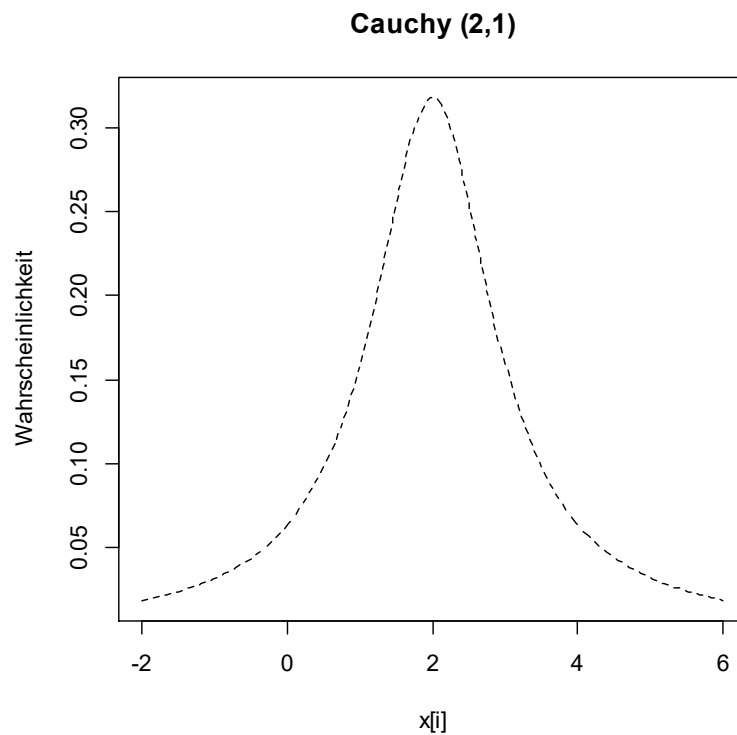


Abb.9: Cauchy-Verteilung mit $x_0 = 2$; $y_0 = 1$ im Intervall $[-2,6]$.

Zum Abschluss der Darstellung der häufigsten Wahrscheinlichkeitsverteilungen, die für die bayesianische Statistik essentiell sind, soll folgende Tabelle eine Übersicht geben, welche konjugierten *priors* zu welcher *Likelihood*-Funktion der Daten passt:

Prior	Likelihood	Posterior
Beta	Binomial	Beta
Beta	Geometrisch	Beta
Gamma	Poisson	Gamma
Gamma	Exponential	Gamma
Normal	Normal (Varianz bekannt)	Normal
Inverse Gamma	Normal (Varianz unbekannt)	Inverse Gamma
Dirichlet	Multinomial	Dirichlet

Tab. 3: Zusammenfassung konjugierter *priors*, angelehnt an Held & Bové, 2014, S. 181.

5.1.3 Likelihood

Am Anfang des Kapitels wurde ein essentieller Teil des Satzes von Bayes beschrieben als die Wahrscheinlichkeit einer Variablen y gegeben eines Parameters θ : $p(y|\theta)$. Zu diesem Term wird in der Fachliteratur meistens ein Äquivalent gebildet, die sog. *Likelihood* des Parameters gegeben einer Variable (die Variable entspricht in der empirischen Forschung den erhobenen Daten). Folgende Formel beschreibt das sog. *Likelihood-Gesetz*:

$$L(\theta|y) \propto p(y|\theta)$$

(vgl. Kaplan, 2014, S. 22). An Hand des *Likelihood-Gesetzes* ist gut erkennbar, dass die oftmals begrifflich singular verwendete *Likelihood* im Wesentlichen eine Funktionenschar des Parameters θ bei unterschiedlichen Konstanten y darstellt. Die Berechnung der *Likelihood* erfolgt vor dem Hintergrund unterschiedlicher Modelle (vgl. ebd., S. 22 ff.). In der frequentistischen Statistik wird bspw. das binomiale Wahrscheinlichkeitsmodell zur *Maximum Likelihood Schätzung* (kurz MLS/MLE) linearer und logistischer Regressionen verwendet. Im Falle eines Testens mehrerer Hypothesen vor dem Hintergrund derselben Daten ist es üblich und hilfreich, den sog. *Likelihood-Quotienten* anzugeben:

$$\frac{L(\theta_1|y)}{L(\theta_2|y)} = \frac{P(y|\theta_1)}{P(y|\theta_2)}$$

(vgl. Arens et al., 2015, S. 1491), wobei θ_i unterschiedliche Werte des unbekannten Parameters vor dem Hintergrund der Daten annimmt y . Eine ähnliche Form dieses Quotienten stellt der sog. *Bayes-Factor* dar, der später in der beispielhaften Anwendung bayesianischer Statistik genau beschrieben wird.

Auch die *Likelihood-Funktion* wird je nach Datensatz in eine mathematische Verteilungsform überführt. Die bereits aufgezeigten Arten unterschiedlicher Wahrscheinlichkeitsverteilungen sind hierfür maßgeblich. Eine mathematische Bestimmtheit der *Likelihood* stellt aus pragmatischer Sicht einen wesentlichen Schritt in der Anwendung bayesianischer Statistik dar, da auf jener Grundlage ggf. die Selektion konjugierter *priors* erfolgt.

5.1.4 Posteriori-Verteilung

Als zentrales Element des Erkenntnisinteresses kann in der bayesianischen Statistik die *Posteriori*-Verteilung $p(\theta|y)$ angesehen werden. Sie entsteht durch die Aktualisierung des transparent gemachten Wissens mit Hilfe empirischer Daten, sprich durch die Kombination der *Likelihood* mit dem *prior*.

In der am Anfang dieses Kapitels dargestellten Form

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

oder mit Hilfe von Proportionalität ausgedrückten Form

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

wirkt die Kalkulation der *Posteriori*-Verteilung recht simpel. Die durch Proportionalität ausgedrückte Form eignet sich jedoch nicht für die exakte Berechnung der *Posteriori*-Verteilung. Die komplett dargestellte Formel dieser benötigt neben allen bereits in diesem Kapitel behandelten Termen den eingangs erwähnten normalisierenden Term (in englischer Literatur häufig als „marginal probability“ beschrieben), der durch den Nenner der rechten Seite gebildet wird. Bei diskretem Parameter θ ergibt sich $p(y)$ aus dem Satz der totalen Wahrscheinlichkeit, der besagt, dass die Einzelwahrscheinlichkeiten **aller** möglichen Ergebnisse j eines Experiments summiert eins ergeben:

$$p(y) = \sum_{i=1}^j p(y|\theta_i)p(\theta_i)$$

(vgl. Arens et al., 2015, S. 1393). Folgendes fiktives Beispiel soll diesen Zusammenhang verdeutlichen:

Sei K das Ereignis Erkrankung an Herpes labiales und $p(K)$ die Prävalenz mit Wert $\frac{3}{10}$. Es sei weiterhin ein Testverfahren mit positivem Testergebnis T zur Bestimmung der Erkrankung vorhanden, das die Sensitivität (Wahrscheinlichkeit, erkrankte Personen korrekt zu erkennen) 95% besitzt

und die Spezifität (Wahrscheinlichkeit, gesunde Personen korrekt zu identifizieren) 80%. Nach dem *Satz der bedingten Wahrscheinlichkeit* ergibt sich die Wahrscheinlichkeit an Herpes labiales erkrankt zu sein unter der Bedingung eines positiven Tests:

$$p(K|T) = \frac{p(T|K)p(K)}{\sum_{i=1}^j p(T|K_i)p(K_i)}$$

$p(T|K)$ ist hierbei die Sensitivität, $p(K)$ die Prävalenz. K_i entspricht allen möglichen Ausprägungen des Merkmals „erkrankt an Herpes labiales“, was für dieses Beispiel den Merkmalen „krank“ K und „nicht krank=gesund“ \bar{K} entspricht und somit $j = 2$ ist. Dadurch ergibt sich

$$p(K|T) = \frac{p(T|K)p(K)}{p(T|K)p(K) + p(T|\bar{K})p(\bar{K})}$$

$p(T|\bar{K})$ entspricht der Wahrscheinlichkeit, einen positiven Test unter der Bedingung nicht krank zu sein zu erhalten. $p(T|\bar{K})$ ergibt sich daher aus $1 - \text{Spezifität}$. $p(\bar{K})$ ergibt sich durch Ausschluss aus $1 - p(K)$. Im konkreten Beispiel ist

$$p(K|T) = \frac{0,95 * 0,3}{0,95 * 0,3 + 0,2 * 0,05} \approx 0,966 \approx 96,6\%$$

Durch die meist uneindeutige Entstehung des Herpes labiales auf Grund unterschiedlicher Dispositionen und verschiedener dafür ursächlichen Herpes simplex-Typen mögen jedoch die Werte der Sensitivität und Spezifität mit einheitlichen und von der Testperson unabhängigen konstanten Ergebnissen nicht zwangsläufig der Realität entsprechen. Ebenso denkbar ist es, dass aus der Belastung mit verschiedenen Herpes simplex Viren keine von der Person unabhängige Wahrscheinlichkeit, überhaupt zu erkranken resultiert.

Handelt es sich also um einen stetigen Parameter θ und um eine ebenfalls stetige Variable y ergibt sich für den normalisierenden Term des Satzes der bedingten Wahrscheinlichkeit im Definitionsbereich $[A,B]$

$$p(y) = \int_A^B p(y|\theta)p(\theta) d\theta$$

(vgl. Kaplan, 2014, S. 14). Äquivalent zum o.gen. Beispiel ergäbe sich

$$p(T) = \int p(T|\theta)p(\theta) d\theta$$

An Hand dieser letzten Gleichung wird deutlich, wie mathematisch komplex die Integration der Funktionen werden kann, je nachdem welcher Art der Funktionen $p(T|\theta)$ und $p(\theta)$ zugehörig sind. Die Berechnung der *Posteriori*-Verteilung komplexer Parameter, die die Berechnung des normalisierenden Terms voraussetzt, wurde aus diesem Grund erst durch die Rechenleistung moderner Computer möglich (vgl. ebd., S. 65).

Die folgenden Abbildungen zeigen kalkulierte *Posteriori*-Verteilungen mit dem R-Package *LearnBayes* (Albert, 2018) zur grafischen Veranschaulichung grundlegender bayesianischer Inferenzstatistik.

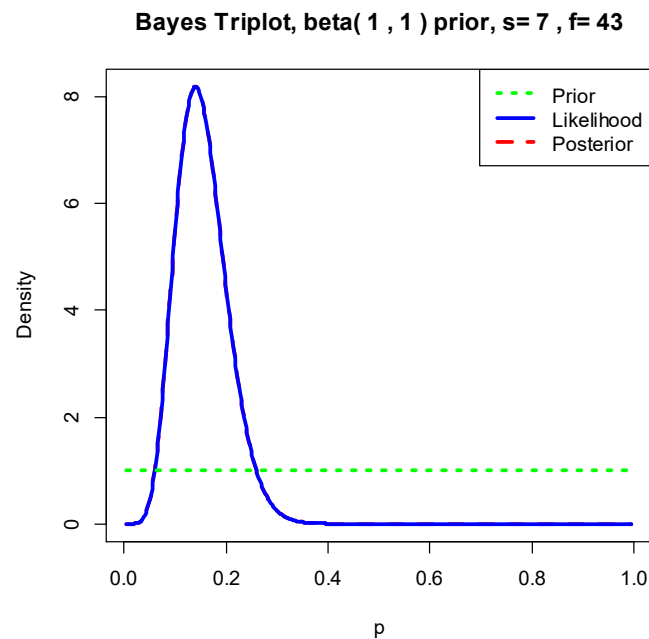


Abb. 10: Annähernd gleichverteilte Beta-Funktion als prior.

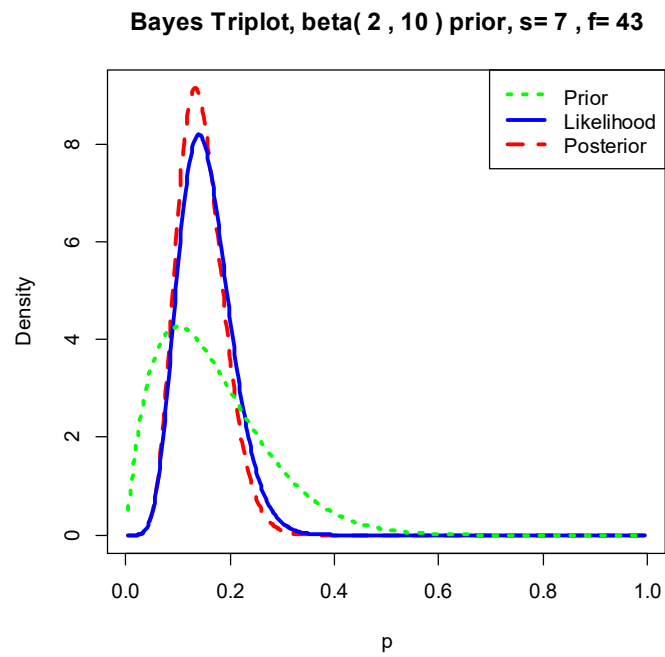


Abb. 11: *Annähernd zentralisierte Beta-Funktion als prior.*

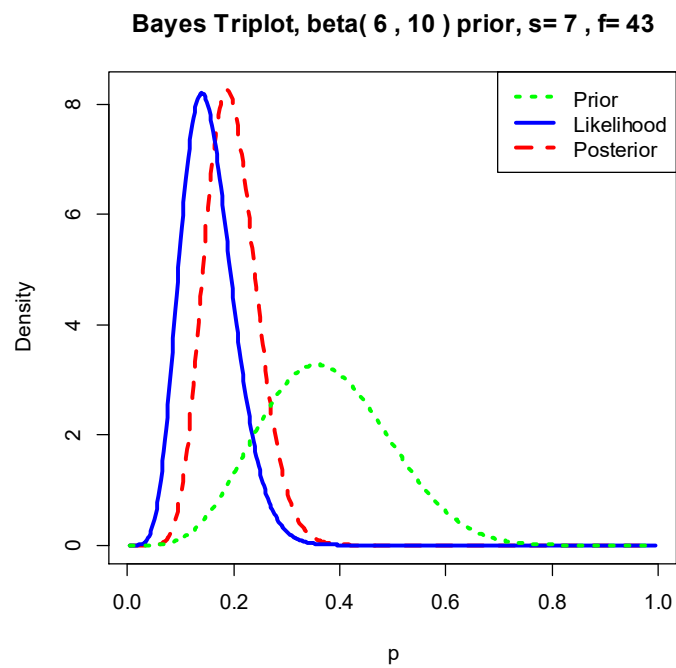


Abb. 12: *Rechtsseitige Beta-Funktion als prior.*

Abbildungen 10 bis 12 zeigen unterschiedliche *Posteriori*-Verteilungen für unterschiedliche *priors* bei gleichbleibender *Likelihood*. Abbildung 10 veranschaulicht den Fall einer Beta-Verteilung die annähernd gleichverteilt ist. Die *Posteriori*-Verteilung ist in diesem Fall identisch mit der *Likelihood*, da der *prior* nicht-informativ ist. Abbildung 11 hingegen illustriert die Verfestigung einer hypothetischen Verteilung, falls *prior* und *Likelihood* in etwa gleiche Wertebereiche als am wahrscheinlichsten beschreiben. In diesem Fall ist das Maximum der *Posteriori*-Verteilung größer als in Abbildungen 10 und 12. In Abbildung 12 ist zu erkennen, dass sich bei einem von der *Likelihood* aus rechtsseitigem *prior* im Gegensatz zu Abbildungen 10 und 11 auch die *Posteriori*-Verteilung rechtsseitig der *Likelihood* befindet. Zusammen illustrieren die Abbildungen gut, welchen Einfluss der *prior* auf das letztendliche Ergebnis hat.

Bevor nun genauer auf die automatisierten Formen bayesianischer Anwendung eingegangen wird, soll als Abschluss der mathematischen Grundlage bayesianischer Statistik die Voraussetzung eben jener Anwendungen eingegangen werden, die ähnlich wie Paradigmen die Struktur statistischer Verfahren im Sinne Bayes bedingen (sollten).

5.2 Voraussetzungen

Das Festlegen von Voraussetzungen für die statistische Praxis sollte keine mathematische Verengung einer grundlegenden Philosophie von Wahrscheinlichkeit darstellen. Viel eher geht es darum, grundlegende Überzeugung in eine analytisch-logische Form zu transformieren, um geeignete statistische Verfahren diskutieren zu können. Hoff (2009) leitet Axiome über sog. ‚Glaubensfunktionen‘ („*belief function*“, S. 14, Hervorhebung im Original) mit Hilfe des Begriffes des rationalen Handelns her, wie er im Kapitel „Historische Betrachtung“ dieser Arbeit dargelegt wird: Eine ‚Glaubensfunktion‘ soll der Bedingung genügen, umso höhere Werte anzunehmen, je höher der Grad der Überzeugung (*belief*) ist „by relating beliefs to preferences over bets“ (ebd.).

Seien $F_G(A)$ und $F_G(B)$ Funktionen, die jeweils eine Menge an Hypothesen (oder eine Menge an Glauben) über ein Phänomen abbilden, soll nach Hoff (ebd.) gelten:

- a. Falls $F_G(A) > F_G(B)$, präferieren wir $F_G(A)$: Das Vertrauen in die erstgenannte Funktion ist größer als in die Zweitgenannte.
- b. Falls $F_G(A|C) > F_G(B|C)$, präferieren wir $F_G(A)$ unter der Voraussetzung, dass das Ereignis C eingetreten ist: Unter der Bedingung von C ist das Vertrauen in erstgenannte Funktion größer als in die Zweitgenannte.
- c. Falls $F_G(A|C) > F_G(A|D)$, präferieren wir $F_G(A)$ eher unter der Bedingung des Ereignisses C als unter der Bedingung von D .

Daraus leitet Hoff (ebd.) ab, dass jede Funktion $F_G()$ folgende Eigenschaften besitzen sollte:

- (I) $F_G(\text{nicht } A|A) \leq F_G(A|B) \leq F_G(A|A)$
- (II) $F_G(B \text{ oder } C|A) \geq \max\{F_G(B|A), F_G(C|A)\}$
- (III) $F_G(B \text{ und } C|A)$ kann bestimmt werden aus $F_G(C|A)$ und $F_G(B|A \text{ und } C)$

Dabei bedeutet (I), dass eine konditionale Glaubensfunktion zwischen einem totalen Misstrauen in die Funktion und einem totalen Vertrauen in die Funktion liegt, (II), dass die Erweiterung einer Glaubensfunktion durch die Verknüpfung ‚oder‘ unter der gleichen Bedingung mindestens den gleichen Wert wie die jeweils einzelnen unverknüpften Funktionen annimmt und schließlich (III), dass aus einer Glaubensfunktion C unter der Bedingung von A und einer Glaubensfunktion B unter Bedingung von C und A eine kombinierte Funktion $F_G(B \text{ und } C)$ unter der Bedingung von A abgeleitet werden kann. Dies kann mit Hilfe dichotomer Glaubensfunktionen mit den Zuweisungen ‚wahr‘ und ‚falsch‘ veranschaulicht werden: Falls C unter der Bedingung von A wahr ist und B unter der Bedingung von $A \text{ und } C$, sind $B \text{ und } C$ unter der einfachen Bedingung von A folglich auch wahr (vgl. ebd., S. 14). Hoff (ebd.) stellt darüber hinaus fest, dass die besagten Eigenschaften mit den klassischen Axiomen der Wahrscheinlichkeit übereinstimmen (vgl. S. 13 ff.).

Für die grundlegende Logik bayesianischer Statistik, die sich in der Kombination von *prior* und *Likelihood* als *Posterior* ergibt, ist der Begriff der ‚Austauschbarkeit‘ notwendig, der in der englischen Fachliteratur als „exchangeability“ (ebd., S. 27) bezeichnet wird. Die Voraussetzung der Austauschbarkeit legitimiert folgende Überlegung: Um ein bestimmtes Modell zur Vorhersage einer Ergebnismenge (der *prior*) zu Grunde zu legen, sind nicht die tatsächlichen unter dieser Bedingung gesammelten Daten notwendig.

Austauschbarkeit an sich meint dabei, dass die Reihenfolge der Daten keine Relevanz für die Bestimmung des Wahrscheinlichkeitsmodells trägt (vgl. Kaplan, 2014, S. 16 f.; Hoff, 2009, S. 27): Sei $p(x_1, x_2, \dots, x_n)$ die Wahrscheinlichkeitsverteilung einer Datenmenge. Falls die Indizes $1, 2, \dots, n$ der Elemente x_1, x_2, \dots, x_n mit Hilfe von π permutiert werden können, ohne dass sich die Wahrscheinlichkeitsverteilung ändert, d.h. falls $p(x_1, x_2, \dots, x_n) = p(x_{\pi 1}, x_{\pi 2}, \dots, x_{\pi n})$, so sind die Elemente x_1, x_2, \dots, x_n austauschbar (vgl. Hoff, 2009, S. 27 f.).

Die Unterscheidung zwischen der Unabhängigkeit von Daten und der Austauschbarkeit von Daten ist äquivalent zur Unterscheidung klassisch frequentistischer und bayesianischer Statistik. Während Erstgenannte zur Modellierung von Daten häufig Unabhängigkeit und identische Verteilung (im Englischen: „independent and identically distributed – often referred to as the *i.i.d. assumption*...“ (Kaplan, 2014, S. 15)) als Voraussetzung fordert (hier erkennt man die Vorstellung von Wahrscheinlichkeit als das rein zufällige Ziehen unabhängiger Kugeln unendlicher Zahl aus einer bodenlosen Urne), fordert eine bayesianische Modellierung lediglich, dass die Reihenfolge der Daten irrelevant für die Datenverteilung ist. Falls Daten dem Anspruch der bedingten Unabhängigkeit genügen, so genügen sie ebenfalls dem Anspruch der Austauschbarkeit (vgl. Hoff, 2009, S. 27 f.). Sind zwei Ergebnisse eines Datensatzes, bzw. einer Wahrscheinlichkeitsverteilung $x_1, x_2 \in p(y)$ bedingt unabhängig voneinander, so gilt: $p(x_1) = p(x_1|x_2)$. Das Element x_2 hat keinen Einfluss auf das Element x_1 . Sollen die Ergebnisse zweier separater Münzwürfe in die Sequenz $p(x_1, x_2)$ zur Beantwortung der Frage, ob beide Münzwürfe das gleiche Ergebnis liefern, überführt werden und es angenommen

werden darf, dass beide Ergebnisse unabhängig voneinander sind, so spielt auch die Reihenfolge der Ergebnisdarstellung (ihre Austauschbarkeit) keine Rolle zur Beantwortung der ursprünglichen Frage.

Jedoch kann aus der Austauschbarkeit von Elementen einer Ergebnismenge nicht direkt auf ihre Unabhängigkeit geschlossen werden. Hoff (ebd.) erklärt mit Hilfe des Rückgriffs auf das sog. ‚de Finetti Theorem‘ (welches besagt, dass aus der Austauschbarkeit einer Daten- oder Ergebnismenge die in dieser Arbeit bereits dargestellte Form des Satzes der bedingten Wahrscheinlichkeit gilt), dass eine Menge an Daten dann dem Kriterium der bedingten Unabhängigkeit genügt, wenn jene austauschbar und wiederholbar ist (vgl. S. 27 ff.). Die Annahme der Wiederholbarkeit stellt sich dabei als eher künstlich und unrealistisch dar, spricht Hoff (ebd.) dabei doch von der Auswahl einer Datenmenge aus einer unendlich großen Grundgesamtheit (und prinzipiell auch der unendlichen Ausprägung jedes einzelnen Ergebnisses) oder der Auswahl einer Datenmenge, bei der jedes Ergebnis in exakter Weise wieder auftreten könnte. Die folgende Abbildung soll den Zusammenhang bedingter Unabhängigkeit, Austauschbarkeit und Wiederholbarkeit verdeutlichen:

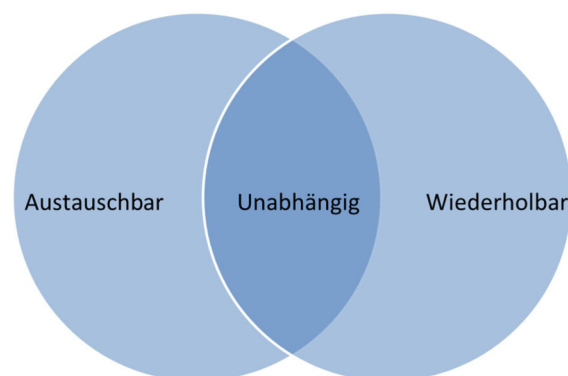


Abb. 13: *Beziehung zwischen mathematischen bayesianischen Voraussetzungen.*

Zusammengefasst stellen sich die logischen Axiome bayesianischer Statistik nicht so weitreichend (oder kritisch ausgedrückt ‚künstlich‘) dar wie

die klassisch frequentistischer Statistik. Sie genügen zwar denen der klassischen Wahrscheinlichkeitstheorie, berücksichtigen jedoch die subjektive Vorstellung Bayes', die in der prägnanten Formel: *Posterior* \propto *Likelihood* * *Prior* mündet. Zum Abschluss des allgemeinen theoretischen Teils dieser Arbeit soll wie bereits angekündigt auf die besonders für ein einführendes Werk interessante und komplexe Form der Kalkulation des normalisierenden Terms oder der *marginal probability* eingegangen werden, ohne die praxisorientierte Fragestellungen schwer oder nicht lösbar wären.

5.3 Die Markow Chain Monte Carlo Methode

Unter dem Begriff ‚Markow Chain Monte Carlo‘ (oder kurz: MCMC) verbirgt sich ein Algorithmus zur Lösung komplexer Integrale von Wahrscheinlichkeitsverteilungen. Die Monte Carlo Methode besteht in der Kalkulation bestimmter *Posteriori*-Werte an Hand zufällig gewählter Produkten von *Likelihood* und *prior*. Die Summierung dieser Werte geteilt durch ihre Anzahl approximiert die tatsächliche *Posteriori*-Verteilung.

Sei die Menge $\{\theta_t, t = 1, 2, \dots, T\}$ eine Auswahl des Parameters θ der *Posteriori*-Verteilung $p(\theta|y)$. Dann ergibt sich diese zu:

$$p(\theta|y) \approx \frac{1}{T} \sum_{t=1}^T p(\theta_t|y)$$

(vgl. Kaplan, 2014, S. 66). Bei der zufälligen Auswahl bestimmter Werte gilt ebenfalls wieder die Voraussetzung der Austauschbarkeit, jedoch nicht der Unabhängigkeit (vgl. ebd., S. 66 f.). Das bedeutet letztendlich, dass die Reihenfolge der Berechnung zufälliger Werte irrelevant für das Ergebnis ist, während der Algorithmus, der die Zufälligkeit simuliert, nicht notwendigerweise für jede Wertkalkulation von neuem beginnen muss. Aus diesem Grund ist der Einsatz sog. Markow-Ketten (Markow Chains) legitim, welche eine Folge von erzeugten Elementen aus einem definierten Werte-, oder Wahrscheinlichkeitsraum bilden mit der Eigenschaft, dass ein erzeugter Wert lediglich durch den vorherigen bedingt ist (vgl. Bättig, 2017,

S. 100 f.). Für eine Folge $f(\theta)$ mit den Werten $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(m)}$ aus dem Wahrscheinlichkeitsraum Θ , für den gilt $\theta^{(i)} \in \Theta, i = 1, 2, \dots, m$, gilt dann: $f(\theta^{(i+1)})$ ist lediglich bedingt durch $f(\theta^{(i)})$. Für die Bildung von Werten die der Art der Markow-Ketten entsprechen stehen verschiedene Algorithmen zur Verfügung. Die in der gesichteten Literatur am Häufigsten erläuterten sind dabei der seit den 1950er Jahren bekannte „*Metropolis-Hastings algorithm*“ (Held & Bové, 2014, S. 270) und das sog. „*Gibbs sampling*“ (ebd.). Die vorliegende Arbeit wird sich auf die Darstellung dieser häufigsten Verfahren beschränken. Held & Bové (2014, vgl. S. 280) weisen ebenso wie Hoff (2009, vgl. S.55) darauf hin, dass der normalisierende Term des Satzes der bedingten Wahrscheinlichkeit auch durch numerische Ansätze gelöst werden kann. Eine Darstellung geht über den Umfang dieser Arbeit jedoch weit hinaus, eine allgemeine Einführung in das Thema kann das aktuelle Werk von Richter & Wick (2017) geben.

5.3.1 Metropolis-Hastings Algorithmus

Zur Bestimmung des zu berechnenden Wertes einer Markow-Kette bedient sich der Metropolis-Hastings Algorithmus einer sog. „*proposal distribution*“ (ebd.), mit deren Hilfe Werte für θ gebildet werden. Diese gebildeten Werte für θ werden anschließend für die zu berechnende *Posteriori*-Verteilung verwendet. Sei die Funktion der *proposal distribution* die *proposal function* $f^*(\theta|\theta^{(m)})$, so wird ein neuer Wert θ^* von dieser Funktion akzeptiert mit der Wahrscheinlichkeit α , für die gilt:

$$\alpha = \min \left\{ 1, \frac{f(\theta^*|x)}{f(\theta^{(m)}|x)} \frac{f^*(\theta^{(m)}|\theta^*)}{f^*(\theta^*|\theta^{(m)})} \right\}$$

(vgl. ebd.). Falls $f(\theta^{(m)}|x) * f^*(\theta^*|\theta^{(m)}) < f(\theta^*|x) * f^*(\theta^{(m)}|\theta^*)$, also falls die Wahrscheinlichkeit des neuen Wertes größer ist als die Wahrscheinlichkeit des alten Wertes, so wird der neue Wert θ^* mit $\alpha = 1$ akzeptiert. Beim gegenteiligen Fall wird der neue Wert hingegen nur mit der Ratio der Wahrscheinlichkeit des neuen und des alten Wertes angenommen (vgl. Kaplan, 2014, S. 68). Held & Bové (2014) stellen fest, dass durch diesen Algorithmus die approximierte Funktion der *Posteriori*-

Verteilung gegen die tatsächliche *Posteriori*-Verteilung konvergiert, unabhängig von der Wahl der *proposal function*. Durch das beschriebene Vorgehen wird jedoch ersichtlich, dass dies nur dann geschieht, wenn die Markow-Kette(n) des Algorithmus' unendlich lang werden. Für nicht annähernd unendliche Sequenzen kann aus der Formel für α abgeleitet werden, dass eine Konvergenz dann schnell erreicht wird, wenn die *proposal function* selbst der tatsächlichen Funktion der *Posteriori*-Verteilung angenähert wird. Der anfängliche Teil der so gebildeten Markow-Ketten werden mitunter „burn-in samples“ (Kaplan, 2014, S. 69) genannt. Damit wird der Teil der Kette beschrieben, der sich nicht in Bereichen hoher Wahrscheinlichkeitsverteilungen befindet und daher (noch) nicht nah an der eigentlich gesuchten Verteilung ist. In diesem gesamten Vorgehen wird exakt die Vorstellung einer subjektiv bedingten Wahrscheinlichkeit sichtbar, die bayesianische Statistik ausmacht. Denn eine optimale Lösung einer MCMC mit einem Metropolis-Hastings Algorithmus wird dann erreicht, falls *prior* und *Likelihood* der Daten perfekt zusammenpassen. Folgende Abbildung zeigt übersichtlich das Vorgehen des Metropolis-Hastings Algorithmus:

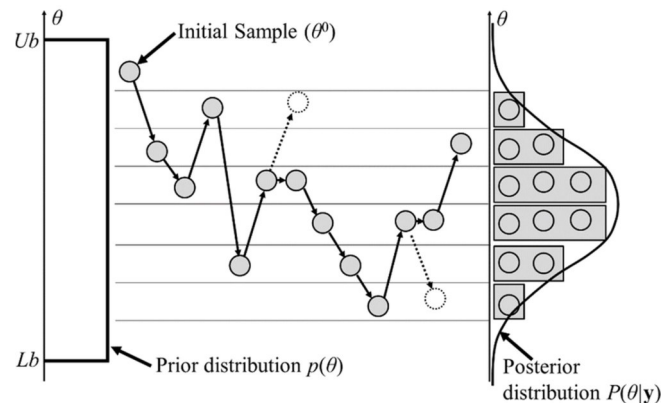


Abb. 14: Schematischer Ablauf des Metropolis-Hastings Algorithmus' (Lee, Sung & Choi, 2015, S. 5543).

5.3.2 Gibbs Sampling

Beinhaltet das gesuchte statistische Modell mehrere Parameter $\theta_1, \theta_2, \dots, \theta_i$ mit der resultierenden *Posteriori*-Verteilung $p(\theta_1, \theta_2, \dots, \theta_i | y)$, wird ein Algorithmus benötigt der mehrere voneinander abhängige Markow-Ketten generiert. Hierfür eignet sich das sog. Gibbs sampling (vgl. ebd.). Bei diesem Verfahren wird die gemeinsame Wahrscheinlichkeitsverteilung der Parameter $\theta_1, \theta_2, \dots, \theta_i$ benötigt. Der Algorithmus beginnt mit einem zufälligen Wert für einen der Parameter, bspw. θ_1^0 . Im Anschluss werden dann Werte aller anderen Parameter unter der Bedingung von θ_1^0 kalkuliert, also $p(\theta_2 | \theta_1^0), p(\theta_3 | \theta_1^0), \dots, p(\theta_i | \theta_1^0)$ mit den daraus folgenden Werten $\theta_2^1, \theta_3^1, \dots, \theta_i^1$. Diese neuen Werte bilden dann Ausgangspunkte für je eine neue Iteration. So werden bspw. für θ_2^1 Werte aller anderen Parameter unter der Bedingung von θ_2^1 gebildet. Formal lässt sich der Prozess folgender Maßen beschreiben (vgl. ebd., S. 69 f.):

Sei θ Vektor von Parametern mit den Elementen $\theta_1, \theta_2, \dots, \theta_i$ und y die gegebene Datenmatrix. Folglich lauten die dazugehörigen Iterationen:

$$1. \text{ Iteration} \quad \theta_1^{(s)} \sim p(\theta_1 | \theta_2^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_i^{(s-1)}, y)$$

$$2. \text{ Iteration} \quad \theta_2^{(s)} \sim p(\theta_2 | \theta_1^{(s)}, \theta_3^{(s-1)}, \dots, \theta_i^{(s-1)}, y)$$

...

$$i. \text{ Iteration} \quad \theta_i^{(s)} \sim p(\theta_i | \theta_1^{(s)}, \theta_2^{(s)}, \dots, \theta_{i-1}^{(s)}, y)$$

mit den dazugehörigen abhängigen Vektoren

$$\theta^{(1)} = \{\theta_1^{(1)}, \dots, \theta_i^{(1)}\}$$

$$\theta^{(2)} = \{\theta_1^{(2)}, \dots, \theta_i^{(2)}\}$$

...

$$\theta^{(s)} = \{\theta_1^{(s)}, \dots, \theta_i^{(s)}\}$$

(vgl. ebd.; Hoff, 2009, S. 96 f.). Der dargestellte Prozess besteht letztendlich aus einer einzigen Markow-Kette. Das Kalkulationsverfahren an sich sieht vor, dass ähnlich wie beim Metropolis-Hastings Algorithmus mehrere solcher Ketten gebildet werden. Die Anzahl an Iterationen i muss ebenfalls vom Benutzer des Algorithmus festgelegt werden.

Mit der exemplarischen Darstellung zweier Algorithmen zur Bestimmung von *Posteriori*-Verteilungen ist der Abschnitt allgemeiner theoretischer und mathematischer Grundlagen für die bayesianische Statistik abgeschlossen. Offenzulegen ist, dass hierbei nur die wichtigsten und essentiellsten Aspekte bayesianischer Statistik aufgezeigt werden konnte. Wie bereits erwähnt, kann bereits die Art und Weise der Auswahl bestimmter *priors* ganze Monographien füllen. Ebenso verhält es sich mit der mathematischen Bestimmung der *Likelihood*. Zu guter Letzt wird auch der jeweilige Algorithmus für die Berechnung der *Posteriori*-Verteilung zu diskutieren sein. Die Darstellung lediglich zweier verschiedener Verfahren lässt dabei die unterschiedlichen Indikationen dieser erahnen. Auch, wenn die mathematische Operationalisierung der Grundlagen bayesianischer Statistik (und erst recht die einzelner bestimmter statistischer Verfahren) mitunter eine gewisse logische Abstraktion und Muße erfordert, ist sie aus Sicht des Autors gerade in dem Wissenschaftsgebiet dieser neuen, größtenteils ungeschriebener Disziplin „Bayesianische Statistik“ besonders wichtig, um ihren Sinn und ihre Denkweise einerseits zu verstehen, andererseits kritisch zu hinterfragen. Eine rein pragmatisch orientierte Denkweise verleitet eher zu einem methodisch unreflektiertem „Trittbrettfahren“. Per definitionem ist der Forscher jedoch angehalten sein phänomenbezogenes Verständnis stets offen zu legen. Im nächsten Abschnitt sollen konkrete bayesianische Verfahren erläutert werden, die sich bisher in der wissenschaftlichen Diskussion herausgebildet haben. Auch hierbei können wiederum nur die Grundzüge dieser Verfahren erläutert werden. Eine etwas ausführlichere Darstellung wird zu dem bayesianischen Äquivalent des klassisch frequentistischen T-Tests gegeben werden, da dieser im darauf folgenden Anwendungsbeispiel zum Einsatz kommt.

5.4 Bayesianische Verfahren

In diesem Abschnitt werden unterschiedliche statistische Verfahren mit bayesianischem Ansatz erläutert. Dazu zählen der sog. Bayes-Faktor, der bayesianische Hypothesentest, das Äquivalent zur linearen Regression und in Ansätzen die Mehrebenenanalyse (kurz: MEA). Dabei kann kein Anspruch auf Vollständigkeit erhoben werden. Weitaus komplexere Methoden als die im Folgenden dargestellten können in weiteren Arbeiten aufgegriffen werden. Die Selektion der Verfahren erfolgte an Hand der am häufigst eingesetzten und bekanntesten Verfahren der Pflege-, bzw. der Sozialwissenschaft (insbesondere das klassische Hypothesentesten und generalisierte lineare Modelle). Dadurch soll einer breiteren Leserschaft Zugang zu dieser Art der Statistik garantiert werden. Im Ausblick wird kurz auf komplexere Verfahren eingegangen.

5.4.1 Bayes-Faktor

Genau genommen ist der Bayes-Faktor kein eigenständiges statistisches Verfahren. Vielmehr kommt er in unterschiedlichen Verfahren als Qualitäts- und Prüfinstrument zum Einsatz. Der Bayes-Faktor vergleicht unterschiedliche Modelle und zeigt auf, wie gut ein bestimmtes Modell im Vergleich zu einem anderen passt (vgl. Cowles, 2013, S. 208). Auf Grund dieser Ambiguität wird der Bayes-Faktor bspw. im Rahmen bayesianischer T-Tests (vgl. Wang & Liu, 2016, S. 195) oder Güteklassifizierung von Regressionsparametern bei der bayesianischen linearen Regression (vgl. Hoff, 2009, S. 164) eingesetzt. Zur Berechnung des Bayes-Faktors werden sog. *odds* gebildet. *Odds* sind Quotienten zweier Wahrscheinlichkeiten. Seien $Pr(M_1)$ und $Pr(M_0)$ *priors* in Form diskreter Wahrscheinlichkeiten und y die dazugehörigen Daten, so ergibt sich der Bayes-Faktor zu:

$$BF_{1,0} = \frac{\frac{Pr(M_1|y)}{Pr(M_0|y)}}{\frac{Pr(M_1)}{Pr(M_0)}}$$

(vgl. Cowles, 2013, S. 209; vgl. Jackman, , S.45). Der Index des Bayes-Faktors gibt an, in welcher Richtung die *odds* gebildet werden. $\frac{Pr(M_1|y)}{Pr(M_0|y)}$ sind dabei die *Posteriori-odds* und $\frac{Pr(M_1)}{Pr(M_0)}$ die *priori-odds*. Die Beachtung der Reihenfolge der getesteten Modelle, bzw. Hypothesen (diese Begriffe werden synonym verwendet) ist für die korrekte Interpretation des Bayes-Faktors notwendig. Der Bayes-Faktor gibt letztendlich an, wie wahrscheinlich ein Modell im Verhältnis zu einem anderen ist. Da sich der Bayes-Faktor folglich als Skalar darstellt, unterliegt seine Interpretation hinsichtlich wissenschaftlicher Relevanz allerdings immer einerseits der Fragestellung und der damit verbundenen Konsequenz (insbesondere im medizinisch-pflegerischen Bereich) für die jeweilige Disziplin und andererseits der Interpretation des Forschers. Folgende Tabelle gibt eine ungefähre Orientierung der Aussagekraft des Bayes-Faktors:

$\log_{10}(B_{10})$	$B_{1,0}$	Evidence against H_0 (or M_0)
0 to 1/2	1 to 3.2	Not worth more than bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	>100	Decisive

Tab. 4: *Tabelle als Hilfe zur Interpretation des Bayes-Faktor (Cowles, 2013, S. 213 n. Kass & Raftery, 1995).*

Die generalisierte Form des Bayes-Faktors für stetige Wahrscheinlichkeiten entspricht dem Quotient der normalisierenden Terme (engl.: *marginal likelihood*) unter Bedingung der jeweiligen Modelle und Parameter, angenommen θ_1, θ_2 :

$$BF_{10} = \frac{\int p(\theta_1)p(y|\theta_1) d\theta_1}{\int p(\theta_2)p(y|\theta_2) d\theta_2} = \frac{p(y|M_1)}{p(y|M_2)}$$

(vgl. Cowles, 2013, S. 211). Wie bereits zuvor thematisiert, erfordert die Integration solcher Terme bspw. MCMC-Verfahren, daher ist die praktische Kalkulation dieser Form des Bayes-Faktors nicht ohne weiteres per Hand möglich. Das bereits erwähnte R-package „BayesFactor“ (Morey & Rouder,

2015) liefert für unterschiedliche Anlässe Berechnungen für den Bayes-Faktor, darunter auch für bayesianische T-Tests und Regressionen.

5.4.2 Bayesianischer Hypothesentest

Bei dem frequentistischen T-Test handelt es sich um einen klassischen Hypothesentest, der mit Hilfe eines Vergleichs zweier unterschiedlicher arithmetischer Mittel durchgeführt wird. Er eignet sich insbesondere zur Unterscheidung zweier Gruppen und geht auf den unter dem Pseudonym „Student“ veröffentlichenden Mathematiker William Sealy Gosset zurück (vgl. Schaefer, 2016, S. 205). Der inflationäre Gebrauch des T-Tests in der empirischen Forschung lässt Gronau, Ly & Wagenmakers (2017) von einem „inferential workhorse“ (S. 3) sprechen. In der Fachliteratur sind mehrere Ansätze als Äquivalenz zum frequentistischen T-Test zu finden. Im Folgenden wird auf die Ansätze von Kaplan (2014), Kruschke (2013) und Gronau et al. (2017) näher eingegangen. In diesem Kontext wird der Begriff von *hierarchischen Modellen* („hierarchical models“ – Cowles (2013, S. 147)) notwendig. Um hierarchische Modelle handelt es sich in der bayesianischen Statistik, falls mehr als ein *prior* und eine *Likelihood* notwendig sind. Dies ist neben Mehrebenenmodellen auch explizit bei dem von Kruschke (2013) beschriebenen Verfahren zutreffend, da hier für verschiedene Parameter (im Kontext von *priors* wird häufig von *hyperparametern* gesprochen) verschiedene *priors* und daraus resultierend mehrere *Likelihoods*, sowie *Posteriori*-Verteilungen modelliert werden.

Die Essenz des bayesianischen T-Tests nach Gronau et al. (ebd.) besteht ebenfalls in der Berechnung eines Bayes-Faktors, jedoch angepasst an die grundlegende Idee des klassischen T-Tests, Mittelwert und Varianz zweier Gruppen zu vergleichen, sowie eine Effektgröße zu integrieren, um daraus einen T-Wert zu erhalten. Die Herleitung und Legitimation dieses Bayes-Faktors, welche insbesondere die Gewichtung des Verhältnisses von *prior* und *Likelihood* (sprich, wie stark ein [großer] Datensatz das Ergebnis beeinflusst) und die Modulation unterschiedlicher *priors* umfasst, stellen Gronau et al. (ebd.) ausführlich dar.

$$BF_{10}(d; \mu_\delta; \gamma; \kappa) = \frac{\int T_\nu(t|\sqrt{n_\delta}\delta) \frac{1}{\gamma} T_\kappa\left(\frac{\delta - \mu_\delta}{\gamma}\right) d\delta}{T_\nu(t)}$$

(vgl. ebd., S. 12) ist dann der resultierende Bayes-Faktor für den T-Test nach ebendiesen. Der Term $T_\nu(t|\sqrt{n_\delta}\delta) = T_\nu(t|a)$ des Zählers stellt eine Dichtefunktion der T-Verteilung dar, mit dem errechneten T-Wert t und dem dezentralisierendem Parameter $\sqrt{n_\delta}\delta = a$, der sich aus der effektiven Stichprobe $n_\delta = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$, mit den Strichproben der beiden Gruppen n_1, n_2 und der standardisierten Effektstärke δ ergibt. Der Term $\frac{1}{\gamma} T_\kappa\left(\frac{\delta - \mu_\delta}{\gamma}\right) = \pi(\delta)$ des Zählers stellt einen t-verteilten *prior* dar, mit dem Streuungsparameter γ und dem Ortsparameter μ_δ unter dem für diese Funktion spezifischen Freiheitsgrad κ . Der Nenner $T_\nu(t)$ mit dem Freiheitsgrad ν stellt sich als standardisierte T-Verteilung mit dezentralisierendem Parameter $a = 0$ dar. Der Unterschied dieses Verfahrens zum klassischen T-Test besteht offensichtlich in der Definition und Modellierung eines funktionalen *priors* $\pi(\delta)$, sowie der wiederum computergestützten Kalkulation des Zählers. An Hand dieser Formel lässt sich die bereits erwähnte Integration spezifischer Bayes-Faktoren in die bayesianische Inferenzstatistik gut illustrieren. Erwähnt werden muss, dass die Wahl und Modulation eines solchen Bayes-Faktors allerdings einem wissenschaftlichen Diskurs unterliegt. Die von Gronau et al. (ebd.) veröffentlichte Version des Bayes-Faktors stellt sich als Resultat verschiedener früherer Arbeiten unterschiedlicher Forscher dar mit dem formulierten Anspruch, berechtigt geäußerte Kritik berücksichtigt zu haben. Insbesondere die von Gronau et al. (ebd.) vorgeschlagene Modellierung des *priors* mit Hilfe einer Dichtefunktion und die Mehrdimensionalität des Ansatzes scheint dem aktuellen wissenschaftlichen Stand angemessen zu sein, wurde die ursprünglich von Gönen et al. (2005) vorgestellte eindimensionale Form des Bayes-Faktors

$$GBF[H_1:H_0](\sigma_a^2) = \left[\frac{1 + \frac{t^2}{\nu}}{1 + \frac{t^2}{\nu(1 + n_\delta\sigma_a^2)}} \right]^{\frac{(\nu+1)}{2}} * (1 + n_\delta\sigma_a^2)^{-\frac{1}{2}}$$

(vgl. Wang & Liu, 2016, S. 195) mit den aus Gronau et al. (2017) bekannten Parametern und mit σ_a^2 als a priori Varianz für T-Tests insbesondere von Wang & Liu (ebd.) doch als zu unflexibel und mathematisch problematisch interpretiert (vgl. S.195 ff.). Die Simplität dieses Ansatzes macht diesen jedoch insbesondere zum Einstieg in die bayesianische Statistik durchaus attraktiv. Die anschließende exemplarische Berechnung eines Bayes-Faktors mit Hilfe des R-packages „BayesFactor“ (Morey & Rouder, 2015) entspricht prinzipiell der hier aufgezeigten Form mit dem Unterschied, dass der gewählte *prior* einer Cauchy- und keiner modifizierbaren T-Verteilung entspricht (vgl. Morey & Rouder, 2011).

Kaplan (2014) schlägt mit Bezugnahme auf den klassischen T-Test eine Kombination aus deskriptiver Analyse der *Posteriori*-Verteilung und die Bildung mehrerer Skalare vor (darunter auch einen Bayes-Faktor). Zur deskriptiven Analyse der *Posteriori*-Verteilung zählt Kaplan (ebd.) das arithmetische Mittel („*expected a posteriori* or EAP“ (S. 94)), die Varianz und das Maximum jener (S. 94 f.). Darüber hinaus stellt Kaplan (ebd.) zwei Intervall-Werte vor: Das „*posterior probability interval*“ (ebd., S. 95) und das „*highest posterior density interval (HPD)*“ (ebd., S. 96). Erstgenanntes gibt die Wahrscheinlichkeit p an, mit der ein Wert θ_x in einem bestimmten Intervall der *Posteriori*-Verteilung liegt. Letztgenanntes entspricht dem Intervall, in dem z -Prozent der höchsten Werte für θ liegen (vgl. S. 95 ff.). Die Differenzierung dieser beiden Werte ist sinnvoll, da bei unsymmetrischen *Posteriori*-Verteilungen bspw. das EAP zu einem anderen Schluss als das HDP führen kann. Die Berechnung eines HPD ist z.B. mit Hilfe des R-packages „HDInterval“ (Meredith & Kruschke, 2018) möglich. Der von Kaplan (2014) angeführte Bayes-Faktor (vgl. S. 102) entspricht der generalisierten Form von Cowles (2013, vgl. S. 211), der in dieser Arbeit bereits dargestellt wurde. Das von Kaplan (2014) erwähnte *Bayesian Information Criterion (BIC)* und das *Deviance Information Criterion (DIC)* (vgl. S. 103 ff.) folgen einer ähnlichen Logik des generalisierten Bayes Faktors. Sie werden an dieser Stelle nur der Vollständigkeit halber erwähnt. Kaplan (ebd.) illustriert damit für den Bereich „Bayesian Hypothesis Testing“ (S. 91) lediglich eine sehr allgemeine Vorgehensweise. Sie steht in Kontrast zu dem bereits

erwähnten detaillierten Verfahren von Gronau et al. (2017). Die Thematisierung deskriptiver Analyse erschien dem Autor jedoch an dieser Stelle äußerst relevant, da sie in Verbindung mit komplexeren Verfahren eine umfassende Darstellung wissenschaftlicher Resultate gewährleistet.

Kruschke (2013) folgert aus der Logik bayesianischer Statistik, dass die Konstruktion einer Alternative zum klassischen T-Test aus der bayesianischen Parameterschätzung an sich entsteht. Damit ginge es nicht primär darum, ein möglichst ähnliches Verfahren des T-Tests im Schein frequentistischer Illumination zu entwerfen, sondern die bayesianische Idee vor dem Hintergrund des Vergleichs zweier Gruppen zu realisieren. Kruschke (ebd.) treibt diese Argumentation auf die Spitze, wenn er behauptet:

„Decisions based on Bayesian parameter estimation are better founded than those based on NHST [null hypothesis significance testing], whether the decisions derived by the two methods agree or not. The conclusion is bold but simple: Bayesian parameter estimation supersedes the NHST *t* test.“ (S. 573).

Kruschke (ebd.) schlägt beim Vergleich zweier Gruppen folgendes Vorgehen vor:

Zu Anfang bedürfe es der Formulierung von *priors* hinsichtlich fünf Parametern: Die arithmetischen Mittel der zwei Gruppen μ_1 und μ_2 , die Standardabweichungen der zwei Gruppen σ_1 und σ_2 , sowie die Normalität („normality“, S. 574) ν der Daten. Kruschke (ebd.) vermeidet im Gegensatz zu Gronau et al. (2017) den Begriff der ‚Freiheitsgrade‘ um die Differenz zum klassisch-frequentistischen Hypothesentesten darzustellen. Der Wert ν kann am einfachsten grafisch interpretiert werden: Je größer ν , desto größer das Maximum der T-Verteilungen und desto schmaler werden die Funktionsränder. Im Grenzwert der T-Verteilung $\lim_{\nu \rightarrow \infty} T_\nu$ approximiert die Funktion an die Normalverteilung. Dabei repräsentieren die a priori Annahmen Eigenschaften zweier T-Funktionen, die Kruschke (ebd.) zur Beschreibung der Daten (und nicht zur Kalkulation von p-Werten wie ein klassischer T-Test) nutzt (vgl. S. 574 f.). Die folgende Abbildung illustriert diese Überlegungen:

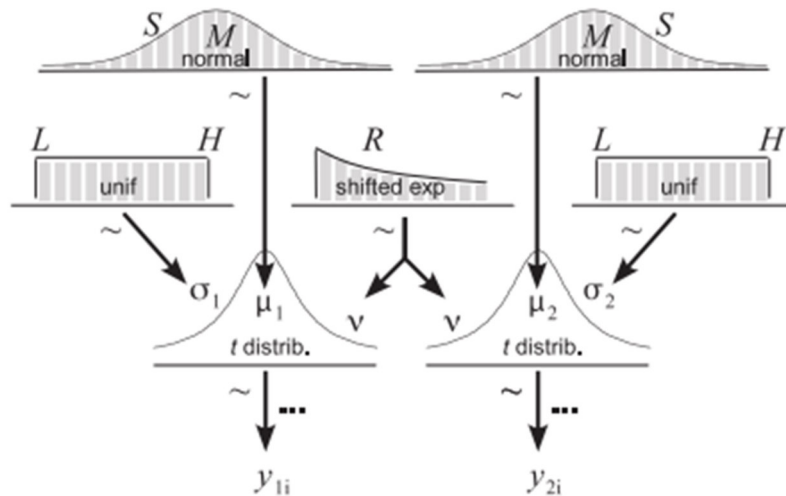


Abb. 15: Hierarchisches Diagramm für bayesianische Parameterschätzung aus Kruschke (2013), S. 575.

Im oberen Teil von Abbildung 15 sind grau schraffierte *priors* erkennbar. Bei den *priors* für die arithmetischen Mittel μ_1 und μ_2 handelt es sich, wie bereits erwähnt, um die Mittelwerte der resultierenden T-Verteilungen. Die *priors* für die Standardabweichungen der zwei Gruppen σ_1 und σ_2 sind durch Gleichverteilungen repräsentiert und stellen so nicht-informative *priors* dar. Der *prior* für die Verteilung der Funktionseigenschaft ν gilt für beide Datensätze (oder Gruppen) und beschreibt die Tendenz zu größeren Werten für ν . Ein größerer Wert für ν hat dabei zur Folge, dass die Ränder der T-Verteilung flacher werden. Werte außerhalb des meist damit verknüpften HDI werden somit unwahrscheinlicher. Kruschke (ebd.) legitimiert diese Vorgehensweise durch die allgemeine Form des Satzes der bedingten Wahrscheinlichkeit

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

mit

$$p(y) = \int p(y|\theta)p(\theta) d\theta$$

(vgl. S. 576). Für die Integration von $p(y)$ wird hierbei auf die Approximation durch MCMC-Verfahren zurückgegriffen (vgl. ebd.). Kruschke (ebd.) verweist darauf, dass die Kalkulation der tatsächlichen *Posteriori*-Verteilung und das in Relation-Setzen zum *prior* und der *Likelihood* eine sicherere Methode darstellt als die bloße Berechnung eines Bayes-Faktors:

„In the Bayes-factor approach, parameter estimates are not emphasized. Moreover, the value of the Bayes factor itself can be very sensitive to the choice of prior distribution in the alternative model.“ (S. 577).

Damit einher geht der Anspruch einer detaillierten Auswertung der Ergebnisse in grafischer und nicht standardisierter Form. Die folgende Abbildung zeigt Beispiele für die Grundlage der Ergebnisinterpretation nach Kruschke (ebd.):

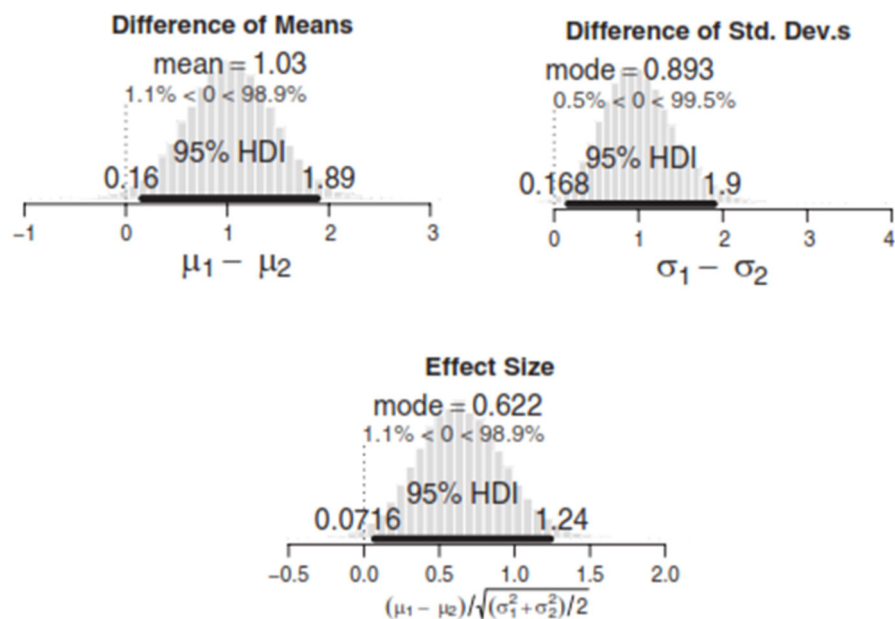


Abb. 16: *Posteriori-Verteilungen unterschiedlicher Kenngrößen nach Kruschke (2013, S. 578).*

Durch die Bestimmung von HDIs kommt Kruschke bei dem in Abbildung 16 gezeigten Beispiels zu dem Schluss eines vorhandenen Effekts. Im Unterschied zum klassischen T-Test wird jedoch an Hand besagter Abbildung deutlich, dass es sich hierbei nicht um absolute Werte im Sinne eines sicher nachweisbaren Effekts handelt, sondern lediglich um eine diskrete Verteilung unterschiedlicher Werte. Vor dem Hintergrund der hierarchischen Struktur, die in Abbildung 15 gezeigt wurde, ist anzunehmen, dass bei diesen resultierenden Verteilungen in erster Linie der Datensatz, bzw. die *Likelihood* der Daten definierend wirkte, wobei die *priors* der arithmetischen Mittel einen symmetrischen *Difference of mean* (Abb. 16, links oben) erst bedingten. Die Interpretation solcher Ergebnisse erschöpft sich also nicht in der Verabsolutierung prägnanter Kenngrößen, sondern stellt eine wesentlich vorsichtiger auf das Wissen und Können des Forschers angewiesene Art der Datenanalyse dar. Kruschke (ebd.) hat das aufgezeigte Vorgehen als *default-setting* in dem R-package „BEST“ (Kruschke & Meredith, 2018) realisiert (vgl. S. 576). Das package wird exemplarisch im Kontrast zum Verfahren von Gronau et al. (2017) am Ende dieses Kapitels an Hand konkreter Daten illustriert.

5.4.3 Bayesianische Lineare Modelle

Die klassisch-frequentistische Idee der Linearen Regression liegt in der funktionalen Beschreibung eines Zusammenhangs mit dem Charakter einer prospektiven Determination (vgl. Schäfer, 2011, S. 85 ff.). Die einfache lineare Regression geht dabei auf die Form

$$y = a + bx = \beta_0 + \beta_i x_i$$

einer Geradengleichung mit b als Steigung und a als Y-Achsenabschnitt zurück. Für multiple lineare Modelle gilt allgemein

$$y = b_0 + b_1 x_1 + b_2 x_2 \dots b_n x_n + \varepsilon_i$$

(vgl. Schäfer, 2011 S. 87) mit den Prädikatoren x_1, x_2, \dots, x_n , den Prädiktor- oder Regressionsgewichten b_1, b_2, \dots, b_n und dem Fehlerterme ε_i . Auch die logistische Regression zählt zu den allgemeinen Linearen

Modellen (kurz: ALM). Durch Gleichsetzen logarithmierter Odds mit der linearen Funktionsgleichung

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i = a + b x_i + \varepsilon_i$$

Ergibt sich eine Wahrscheinlichkeitsfunktion

$$p_i = \frac{1}{1 + e^{-(a+b x_i + \varepsilon_i)}}$$

mit den bereits bekannten Parametern und Koeffizienten. Allerdings entspricht e hierbei der Euler'schen Zahl.

Im Kontext der einfachen linearen Regression stellt Cowles (2013) fest, dass für eine bayesianische Vorgehensweise im Falle einer einfachen Linearen Regression der Form $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ das hauptsächliche Interesse auf der Bestimmung des Regressionskoeffizienten β_1 liegt, wobei sich die unbekannten Parameter zu der Konstanten β_0 , dem Regressionskoeffizienten β_1 und der Varianz σ^2 der normalverteilten³ Fehler ε_i (vgl. S. 179 f.) ergeben. Aus dieser Überlegung heraus ergibt sich die normalverteilte *Likelihood* der Daten zu:

$$y_i | x_i, \beta_0, \beta_1, \sigma^2 \sim N(\beta_0 + \beta_1(x_i - \bar{x}), \sigma^2), i = 1, \dots, n$$

(vgl. ebd., S. 187). Für **nicht-informative priors**, also gleichverteilte oder annähernd gleichverteilte Parameter $\beta_0, \beta_1, \sigma^2$ ergibt sich die gesuchte *Posteriori*-Verteilung von β_1 als T-Verteilung

$$\beta_1 | y \sim t\left(\hat{\beta}_1, \frac{s^2}{\sum(x_i - \bar{x})^2}, n - 2\right)$$

(vgl. ebd., S. 189). Wobei der durch die *Likelihood* bestimmte Regressionskoeffizient $\hat{\beta}_1$ das arithmetische Mittel der T-Verteilung ist, sowie der aus Regressionsvarianz s^2 und der quadrierten Summe aller zentrierten x -Werte $\sum(x_i - \bar{x})^2$ kombinierte Streuungsparameter $\frac{s^2}{\sum(x_i - \bar{x})^2}$ der T-Verteilung und die Freiheitsgrade der T-Verteilung durch $n - 2$ definiert

³ Dies ist eine Annahme, die Cowles aus der frequentistischen Form der Linearen Regression übernimmt

werden (vgl. ebd. S. 187 ff.). Äquivalent dazu ist die Konstante β_0 beschreibbar durch

$$\beta_0|y \sim t\left(\hat{\beta}_0, \frac{s^2}{n}, n-2\right)$$

und die Varianz als inverse Gammaverteilung durch

$$\sigma^2|y \sim IG\left(\frac{n-2}{2}; \frac{SSR}{2}\right)$$

(vgl. ebd., S. 188). *SSR* bezeichnet hierbei „sum of squared residuals“ (ebd., S. 181), ein Maß der klassischen Linearen Regression für die Abweichung von vorhergesagten und tatsächlichen Werten. Cowles (ebd.) weist daraufhin, dass für *Posteriori*-Verteilungen, die die bereits erwähnten Voraussetzungen erfüllen sollen insbesondere $n > 2$ gilt (vgl. S. 189).

Für **informative priors** stellt Cowles (ebd.) lediglich fest, dass auf Grund gängiger Vorannahmen über die Verteilungsformen von $\beta_0, \beta_1, \sigma^2$ (Konstante und Regressionskoeffizienten seien (multivariat) normalverteilt und die Varianz folge einer inversen Gamma-Verteilung) mit keinen konjugierten *priors* für die gemeinsame Wahrscheinlichkeitsverteilung von $\beta_0, \beta_1, \sigma^2$ kalkuliert werden könne. Eine allgemeine mathematische Form wird dabei nicht präsentiert (vgl. S. 192). Auch Hoff (2009) verweist in diesem Zusammenhang auf den Rückgriff eines Gibbs-samplers (vgl. S. 155), also auf eine MCMC-gestützte Approximation der gemeinsamen *Posteriori*-Verteilung. Kaplans (2014) äußerst knappe Darstellung bayesianischer Linearer Regression und der Darstellung einer gemeinsamen Wahrscheinlichkeitsverteilung für $\beta_0, \beta_1, \sigma^2$ im speziellen ist letztendlich ebenfalls auf eine computergestützte Berechnung angewiesen und verlangt ein großes Maß an Abstraktionsvermögen (vgl. S. 125). Der Autor hält eine Abbildung in dieser Arbeit aus diesem Grund für widersinnig. Der Logik dieses Verfahrens folgen unterschiedliche Programme. Für das Statistikprogramm R eignet sich bspw. das package „BLR“ (los Campos & Rodriguez, 2014). Die Essenz bayesianischer Linearer Regression und in ihrer Konsequenz auch die Essenz generalisierter linearer Modelle, wie bspw. der logistischen Regression, stellt sich letztendlich in einer komplexeren Analyse von Verteilungen der

Konstanten, Regressionskoeffizienten und der Fehlervarianz dar. Der große Unterschied zu frequentistischen Vorgehensweisen liegt auch hier in der Transparenz etwaigen Vorwissens und der Analyse der durch Kombination mit den Daten approximierten *Posteriori*-Verteilung(en). Die so kalkulierten Ergebnisse unterliegen natürlicher Weise nicht dem Anschein eindimensionaler, diskreter Erkenntnis. Die folgende Abbildung soll dies verdeutlichen:

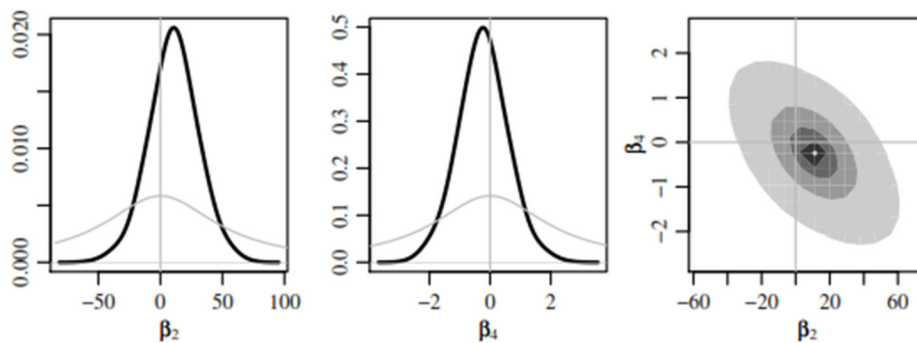


Abb. 17: *Posteriori-Verteilungen zweier Regressionskoeffizienten aus Hoff (2009), S. 160.*

Statt mit Signifikanzniveaus versehene skalare Regressionskoeffizienten β_2, β_4 zeigt die linke und die mittlere Grafik von Abbildung 17 Verteilungsfunktionen, mit relativ 0-zentriertem Maximum. Die gemeinsame *Posteriori*-Verteilung rechts davon verdeutlicht in der Gesamtheit der Abbildung 17 die aus dieser Kalkulation mittransportierte Vagheit der Ergebnisse durch die relativ hohe Ausdehnung des Wahrscheinlichkeitsraums im Verhältnis zur Fläche der maximalen Wahrscheinlichkeit. So interpretiert tendiert eine bayesianische Analyse also nicht zu einer Absolutierung verschiedener diskreter Werte, sondern zu einer Relativierung und zu einer aufwändigeren, vorsichtigeren Interpretationsweise. Äquivalent zu der Anwendung generalisierter linearer Modelle stellt sich eine bayesianisch motivierte Mehrebenenanalyse dar. Diese benötigt jedoch neben den in diesem Abschnitt beschriebenen ein weiteres Prinzip differenzierter Analyse.

5.4.4 Bayesianische Mehrebenenanalyse

Das Konzept einer bayesianischen Mehrebenenanalyse ergibt sich prinzipiell durch die Kombination der bereits im Kontext bayesianischer Hypothesentests erwähnten hierarchischen Modellierung und der Anwendung der generalisierten Linearen Regression [kurz: GLA] (vgl. Hoff, 2009, S. 195). Die klassisch frequentistische Mehrebenenanalyse (kurz: MEA, engl. Multilevel Analysis [kurz: MLA]) berücksichtigt Zusammenhänge unterschiedlicher Dimensionen und verhindert so Fehlschlüsse, die auf der Annahme der Unabhängigkeit von Daten beruhen, wie bspw. der Über- oder Unterschätzung aggregierter Daten (vgl. bspw. Hox, 2010). Es geht dabei um die Analyse von Zusammenhängen innerhalb von Gruppen und zwischen Gruppen. Für ein zweidimensionales MEA-Modell ergibt sich die Regressionsgleichung bspw. zu

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1i} + \beta_{2j} X_{2i} + e_{ij}$$

wobei X_{1ij} und X_{2ij} die Variablen auf individueller Ebene darstellen, e_{ij} die Residuen und β_{0j} , β_{1j} , β_{2j} die Regressionskoeffizienten. Wie durch die Indizes ersichtlich, variieren die Regressionskoeffizienten gruppenabhängig. Hoff (2009) erklärt das allgemeine Vorgehen hierarchischer bayesianischer GLR als die Erklärung von Heterogenität innerhalb der Gruppen durch die bereits erwähnte GLR und die Erklärung von Varianz zwischen den Gruppen mit Hilfe eines weiteren bayesianischen Wahrscheinlichkeitsmodells (vgl. S. 196). Das Modell für die Heterogenität innerhalb der Gruppen wird repräsentiert durch die Gleichung

$$Y_{i,j} = \beta_j^T x_{i,j} + \epsilon_{i,j}, \{\epsilon_{i,j}\} \sim i.i.d. normal(0, \sigma^2)$$

(vgl. ebd.), wobei $x_{i,j}$ der eindimensionale ($p \times 1$) Vektor der Regressoren für das Datum i in der j -ten Gruppe ist, β_j^T die transponierte Matrix der Regressionskoeffizienten für j Gruppen ist, $\epsilon_{i,j}$ den Fehlerterm darstellt und $Y_{i,j}$ den Wert der abhängigen Variable Y des i -ten Datum der j -ten Gruppe. Der Ausdruck $i.i.d. normal(0, \sigma^2)$ verweist auf die Annahme, dass die Residuen $\epsilon_{i,j}$ unter dem $i.i.d.$ Kriterium (independent and

identically distributed random variables) normalverteilt sind mit dem Mittelwert 0 und der Varianz σ^2 . Dies kann nach Hoff (2009) äquivalent ausgedrückt werden als

$$Y_j \sim \text{multivariat normal}(X_j \beta_j, \sigma^2 I)$$

(vgl. ebd.), wobei Y_j ein gruppenspezifischer Datenvektor der Werte $Y_{1,j}, \dots, Y_{n_j,j}$ ist, X_j eine $n_j \times p$ -Matrix der Regressoren $x_{1,j}, \dots, x_{n_j,j}$, also (X_1, \dots, X_m) darstellt, β_j die Regressionskoeffizienten β_1, \dots, β_m und I die $p \times p$ Einheitsmatrix.

Die Heterogenität zwischen den Regressionskoeffizienten wird dann wiederum mit einem eigenen Wahrscheinlichkeitsmodell abgebildet, das durch die Approximation

$$\beta_1, \dots, \beta_m \sim i.i.d. \text{multivariat normal}(\theta, \Sigma)$$

beschrieben wird (vgl. ebd., S. 197). Grafisch lässt sich diese Überlegung mit Hilfe der folgenden Abbildung verdeutlichen:

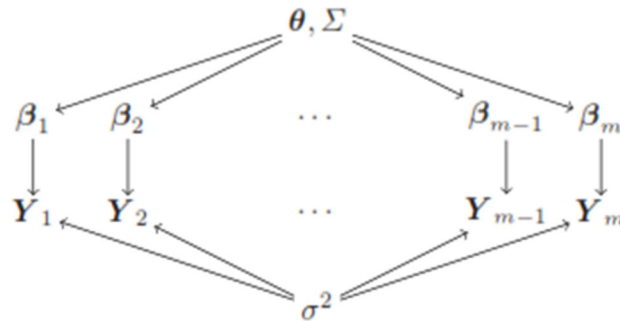


Abb. 18: Hierarchisches Modell zur Bestimmung der Heterogenität zwischen den Gruppen einer mehrdimensionalen bayesianischen GLR.

Aus Abbildung 18 geht hervor, dass für die *Posteriori*-Verteilung von β_j die Festlegung von *priors* für die *hyperparameter* θ, Σ, σ^2 notwendig ist. Unter der Annahme verschiedener Verteilungen für die *hyperparameter* kann die *Posteriori*-Verteilung durch die Form

$$p(\beta_1, \dots, \beta_m, \theta, \Sigma, \sigma^2 | X_1, \dots, X_m, y_1, \dots, y_m)$$

(vgl. ebd., S. 198) bestimmt werden. Dafür sind wiederum MCMC-gestützte Kalkulation der gängige Weg. Folgende Abbildung zeigt eine *Posteriori*-Verteilung eines variablen Regressionskoeffizienten (*slope*) θ_2 :

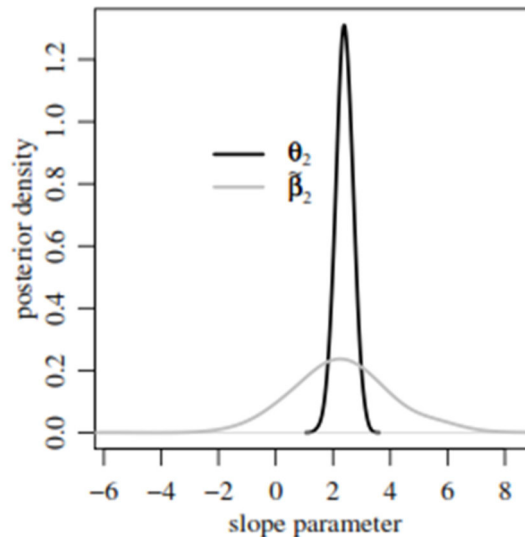


Abb. 19: *Posteriori-Verteilung des slope Parameters θ_2 in schwarz aus Hoff (2009), S. 200.*

Durch Abbildung 19 wird äquivalent zu der Beschreibung eindimensionaler GLR deutlich, dass die verteilungsbasierte Analyse gesuchter Parameter bei dem bayesianischen Pendant der MEA nicht auf eindeutiger Interpretation diverser funktionaler Zusammenhänge beruht. Durch die ohnehin komplexe Natur bayesianischer Parameterschätzung ergeben sich darüber hinaus weitere nicht triviale Probleme, wie bspw. die überaus anspruchsvolle Selektion von *priors* in mehrdimensionalen Modellen (vgl. Samaniego, 2010, S. 122). Auf einzelne spezifische Aspekte zur Lösung dieser Probleme kann jedoch nicht in dieser Arbeit eingegangen werden. Mit dieser bereits weit fortgeschrittenen Analysemethode soll die Illustration theoretisch-mathematischer Operationalisierung bayesianischer Statistik abgeschlossen werden. Im folgenden Anwendungsbeispiel wird das Prinzip des Bayes-Faktors und dem bayesianischen Hypothesentestens noch einmal dargestellt, bevor die inhaltliche Bedeutung bayesianischer Statistik für die Pflegewissenschaft und –forschung thematisiert wird.

5.4.5 Anwendungsbeispiel: Bayes-Faktor und bayesianisches Hypothesentesten

In diesem Abschnitt soll mit Hilfe der R-packages „BayesFactor“ (Morey & Rouder, 2015) und „BEST“ (Kruschke & Meredith, 2018) eine exemplarische Re-Analyse bereits ausgewerteter Daten erfolgen, um die Differenzen klassisch frequentistischer und bayesianischer Statistikverfahren zu demonstrieren. Im Zuge der sog. *good clinical practice* und des *evidence based nursing* kann eine Orientierung an und Übertragung von klassischen medizinischen Interventionsstudien in den Bereich der Pflegeforschung exploriert werden (vgl. Kent & McCormack, 2011, S. 99 ff.). Der in diesem Feld weit verbreitete Einsatz von T-Tests gibt dieser exemplarischen Darstellung daher besondere Relevanz durch das Aufzeigen alternativer Analysemethoden.

In einer im Jahr 2006 erschienenen Interventionsstudie von Dauer, Kelvin, Horan & Germain wurde eine Gruppe von 113 onkologischen *registered nurses*⁴ in New York City hinsichtlich ihres fachlichen Wissens zu Sicherheitsaspekten bei der onkologischen Strahlentherapie untersucht. Ein *pretest-posttest* Design sah keine Kontrollgruppe vor. Die Pflegefachkräfte wurden unter anderem einem schriftlichen Test vor und nach einem umfänglichen Sicherheitstraining unterzogen. Es handelte sich dabei um einen Multiple Choice Test mit vier Antwortmöglichkeiten für jedes der 15 Items. Auf dieser Grundlage wurde ein klassischer T-Test durchgeführt. Die gegebenen Antworten wurden mit Hilfe von Anteilsbildung (Anteil richtiger Antworten geteilt durch gesamte Antworten) zu metrischen Variablen transformiert. Diese Transformation auf Grundlage der dichotomen und interpretatorisch uneindeutigen Zuweisung von wahr/falsch kann dabei kritisiert werden. Außerdem muss die Interventionswirksamkeit ohne eine Vergleichsgruppe ebenfalls kritisch betrachtet werden, da bereits die Wiederholung eines identischen Tests zu einem verbesserten Gesamtergebnis führen könnte. Die folgende Tabelle zeigt eine Zusammenfassung deskriptiver Werte der Studie:

⁴ *Registered nurses* durchlaufen eine mehrjährige Ausbildung und schließen diese im Normalfall mit einer staatlichen Prüfung und einem Bachelorgrad ab.

Descriptive Statistic	Pretest	Posttest
Number of Tests Graded	113	113
Number of Graded Items	15	15
Total Score Possible	100.0%	100.0%
Maximum Score	86.7%	100%
Minimum Score	20.0%	20.0%
Median Score	60.0%	73.3%
Mean Score	58.9%	71.2%
Variance	157.1	386.2
Standard Deviation	12.5	19.7

Tab. 5: Kennzahlen der Studie von Dauer et al. (2006), S. 4.

Der mit Hilfe eines T-Tests berechnete T-Wert wird von Dauer et al. als hoch signifikant bewertet. Vor dem Hintergrund der Forschungsfrage wird die Wirksamkeit der Intervention daher als sehr groß interpretiert. Folgende Tabelle zeigt eine Zusammenfassung der T-Statistik:

Table 2: t-test: paired two sample for means of the cognitive pretest versus posttest

	Pretest	Posttest	t-test
Number of Observations	113	113	
Mean	58.9%	71.6%	
Variance	157.1	385.7	
Degrees of freedom	112	112	224
Calculated t-value			5.56
Critical t-value (two-tail, alpha = 0.05)			1.98
p-value			1.85×10^{-7}

Tab. 6: Kennwerte der T-Statistik aus der Studie von Dauer et al. (2006), S. 8.

Die Angabe der Freiheitsgrade (*Degrees of freedom*) in der Spalte „t-test“ mit einem Wert von 224 legt die Vermutung nahe, dass Dauer et al. zur Berechnung einen T-Test mit zwei unabhängigen Stichproben genutzt haben. Dies widerspricht dem Design eines *paired two sample t-test*, wie die Überschrift der Tabelle im Original angibt. Darüber hinaus ist anzumerken, dass die Mittelwerte und die Varianz der Posttest-Gruppe in Tabelle 5 und Tabelle 6 voneinander abweichen. Eine Rekalkulation der T-Statistik mit R ergab folgendes Ergebnis:

```

> x1<-rnorm(mean=58.9,sd=12.5,n=113)
> x2<-rnorm(mean=71.6,sd=19.6,n=113)
> t.test(x=x1, y=x2, paired=TRUE, alternative="two.sided",var.equal = FALSE)

    Paired t-test

data:  x1 and x2
t = -5.3311, df = 112, p-value = 5.121e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -16.344417  -7.487149
sample estimates:
mean of the differences
      -11.91578

```

Wie an Hand des Ergebnisses zu sehen ist, unterscheidet sich die Rekalkulation des T-Wertes lediglich gering von der ursprünglich von Dauer et al. (2006) vorgenommenen Berechnung. Mit einem T-Wert von ca. 5,33 und einer Wahrscheinlichkeit der Nullhypothese, die gegen null geht, kann auch in diesem Fall von einem äußerst signifikanten und in der Logik einer frequentistischen Statistik von einem äußerst verlässlichen Ergebnis gesprochen werden. Nachdem durch diese erneute Berechnung des T-Wertes von einem ungefähr korrekten T-Wert ausgegangen werden kann, soll der Bayes-Faktor n . Morey & Rouder (2015) berechnet werden. Zur Berechnung des Bayes-Factors werden beide errechneten T-Werte verwendet. Die notwendige Annahme für den Streuungsparameter der Cauchy-Verteilung, „rscale“, welche als *prior* für die Effektstärke fungiert (vgl. ebd.), wird auf einen Wert von ca. $\frac{\sqrt{2}}{2} \approx 0,7011$ gesetzt. Dies entspricht einer eher engeren Cauchy-Verteilung und wird auf Grund der von Dauer et al. (2006) erwähnten Annahme gewählt, dass die Intervention eine hohe Wirksamkeit besitzt (vgl. S. 2). Die Berechnung mit Hilfe von R liefert folgendes Ergebnis (das Vorzeichen des T-Werts spielt keine Rolle):

```

> ttest.tstat(t=5.56,n1=113,rscale=0.70711,simple=TRUE)
      B10
65860.8

> ttest.tstat(t=5.33,n1=113,rscale=0.70711,simple=TRUE)
      B10
24947.95

```

Unabhängig vom verwendeten T-Wert sprechen die enorm hohen Bayes-Faktoren von ca. 65861 für einen T-Wert von 5,56 und 24948 für einen T-Wert von 5,33 für ein sehr bedeutendes Ergebnis (vgl. Cowles, 2013, S. 213). In der Folge wäre auch aus Sicht des Bayes-Faktors die Alternativhypothese des T-Tests absolut klar zu favorisieren.

Mit Hilfe des R-packages „BEST“ (Kruschke & Meredith, 2018) können *Posteriori*-Verteilungen anhand eines MCMC-Verfahrens approximiert werden. Aus der Studie von Dauer et al. (2006) sind approximative Primärdaten des Multiple Choice Tests aus den folgenden Abbildungen ableitbar⁵:

⁵ Die Bitte an Herrn Lawrence T. Dauer um Zusendung der Primärdaten wurde vom Autor gestellt. Eine Antwort zur rechtzeitigen Einarbeitung erhielt der Autor jedoch nicht. Sollte eine Veröffentlichung aus der vorliegenden Arbeit entstehen, wird das BEST-Verfahren mit Hilfe der Primärdaten neu kalkuliert. Zur exemplarischen Illustration spielen die genauen Daten jedoch nur eine untergeordnete Rolle.

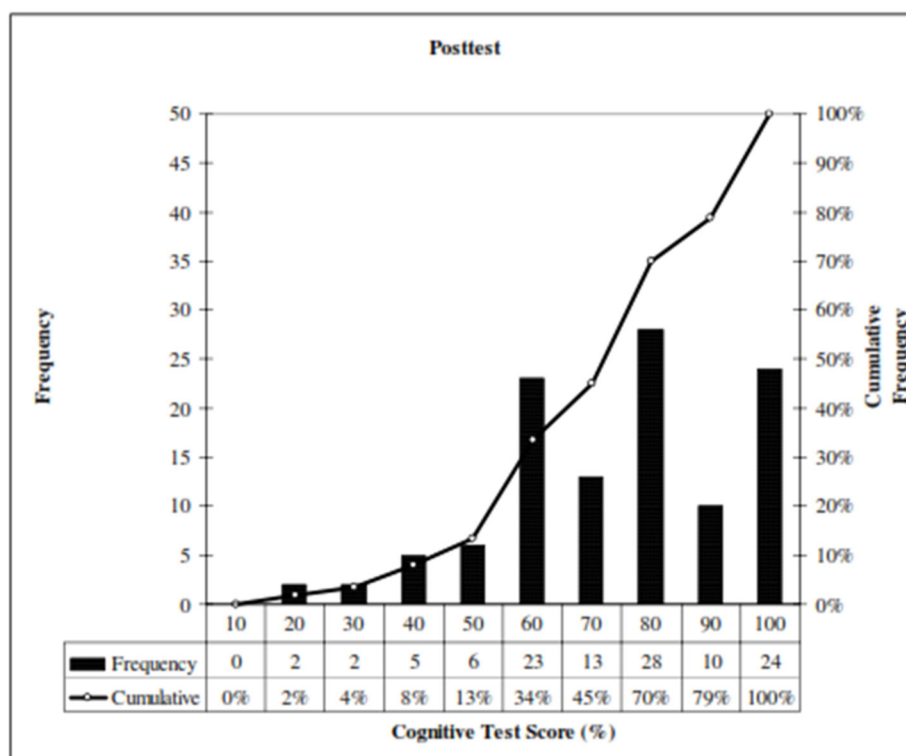
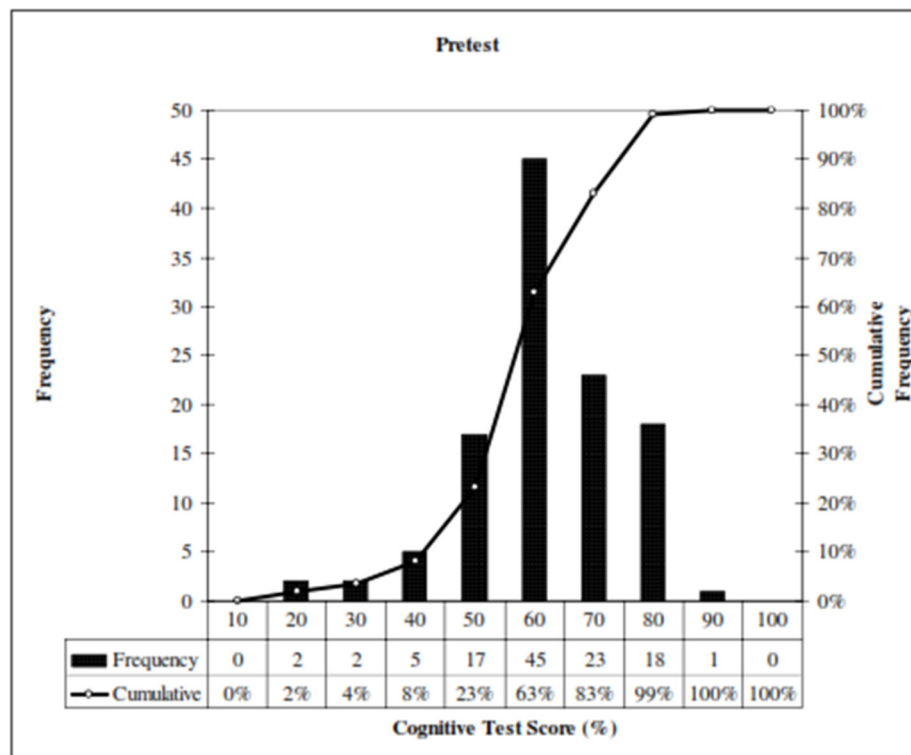


Abb. 20: Histogramm-artige Abbildungen der Testergebnisse aus der Studie von Dauer et al. (2006), S. 5 f. .

Aus den in Abbildung 20 dargestellten Histogrammen kann durch Übertrag der frequenziellen 10%, 20%, usw. -Perzentilen eine Datenmatrix für beide Zeitpunkte der Messung erstellt werden. Eine Zusammenfassung der beiden Vektoren (y1=Pretest-Gruppe; y2=Posttest-Gruppe) dieser Matrix zeigt die Abweichung zu den tatsächlichen Daten:

```
> numSummary(y1)
      mean      sd IQR 0% 25% 50% 75% 100%   n
61.9469 13.15148  10 20  60  60  70  100 113

> numSummary(y2)
      mean      sd IQR 0% 25% 50% 75% 100%   n
74.60177 19.68663  30 20  60  80  90  100 113
```

Der Mittelwert für die Pretest-Gruppe weicht dabei um ca. 5% ab, ebenso wie die Standardabweichung und der Mittelwert in der Posttest-Gruppe. Die approximierte Standardabweichung der Posttestgruppe ist nahezu identisch zu der tatsächlichen. Das im vorherigen Abschnitt dargestellte Verfahren nach Kruschke (2013), das mit Hilfe der Funktion „BESTmcmc“ des R-packages „BEST“ realisiert wurde, liefert folgendes Ergebnis:

```
MCMC fit results for BEST analysis:
100002 simulations saved.
      mean      sd median HDIlo HDIup  Rhat n.eff
mu1   62.18  1.226  62.18 59.775 64.58 1.000 60767
mu2   75.03  1.898  75.04 71.329 78.78 1.000 58875
nu    29.69 25.029  21.68  4.091 80.00 1.001 12185
sigma1 12.35  1.072  12.33 10.219 14.42 1.000 28510
sigma2 19.06  1.471  19.00 16.181 21.96 1.000 40676

'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
'n.eff' is a crude measure of effective sample size.
```

Dabei ist zu beachten, dass auf Grund der sehr vagen und intransparenten Explikation des bereits vorhandenen Wissens von Dauer et al. (2006) keine weitere Modifikation der *priors* der beiden Mittelwerte, Standardabweichungen und der Normalisierung vorgenommen werden konnte. Die default-settings der Funktion „BESTmcmc“ entsprechen daher der im vorherigen Abschnitt von Kruschke (2013) diskutierten. Darüber hinaus wurden die burn-in Schritte auf 1000 festgelegt und die Anzahl der

Observationen auf $1 * 10^5$, was einem gängigen Procedere entspricht (vgl. Kruschke & Meredith, 2018). Nichtsdestotrotz liefert der Output der approximierten *Posteriori*-Verteilungen wertvolle Informationen: Erkennbar ist, dass sich die 95% HDI credible intervals der Mittelwerte nicht überschneiden ($HDI_{lo}(\mu_2) > HDI_{up}(\mu_1)$). Ebenso ist dies bei den Standardabweichungen beider Gruppen festzustellen ($HDI_{up}(\sigma_1) < HDI_{lo}(\sigma_2)$). Ein Unterschied der beiden Mittelwerte kann daher als gegeben interpretiert werden, ebenso eine weitaus größere Varianz in der zweiten Gruppe. Mit Hilfe der Funktion „plot.BEST“ des packages „BEST“ können die Ergebnisse auch in grafischer Weise dargestellt werden:

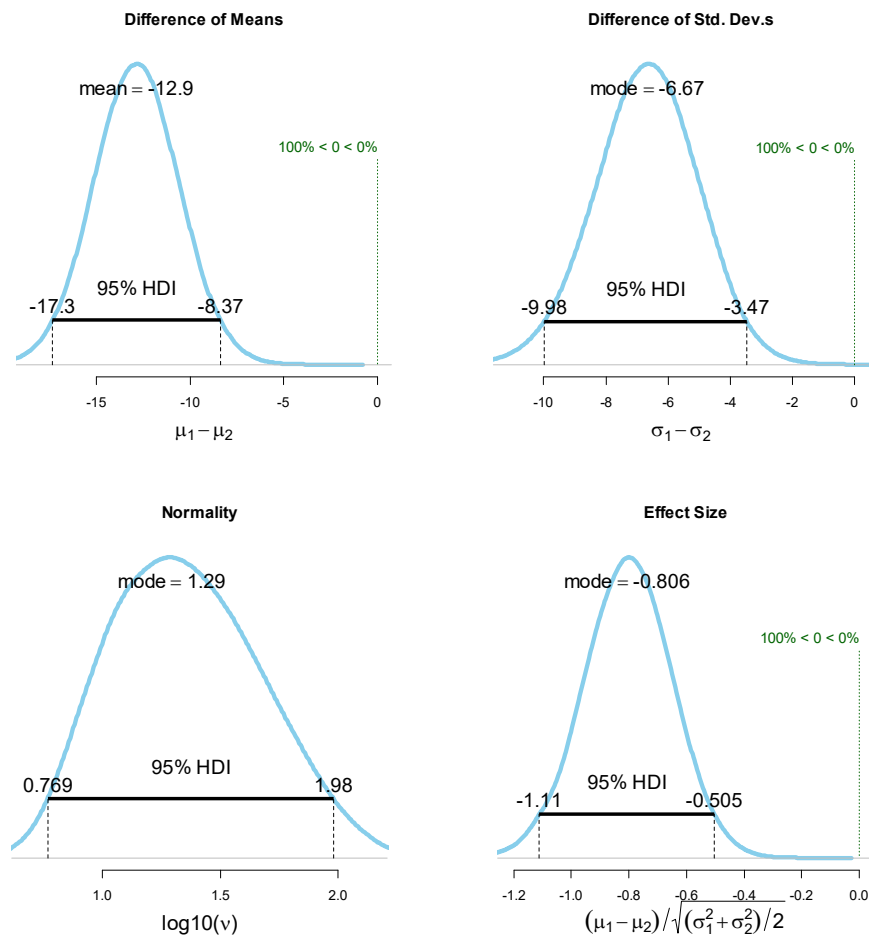


Abb. 21: Ergebnisse der MCMC-gestützten Approximation von *Posteriori*-Verteilungen unterschiedlicher Parameter.

Die Abbildung 21 zeigt die *Posteriori*-Verteilungen der Mittelwertsdifferenzen (links oben), Differenzen der Standardabweichungen (rechts oben), der logarithmierten Normalisierung und der Effektstärke. Auch hierbei ist ein bestehender Effekt zwischen den Gruppen nicht zu leugnen, wie die Graphen der Mittelwertsdifferenz und der Effektstärke abbilden. Allerdings zeigen diese Graphen auch, dass das Ausmaß des Effekts der Intervention nicht zwangsläufig als so sicher und überragend anzunehmen ist, wie der klassische T-Test vermuten ließe. Das 95% HDI der Effektstärke reicht, wie zu erkennen ist, schließlich von einem mittleren Effekt (Betrag=0,505) bis zu einem großen Effekt (Betrag>0,8), die Mittelwertsdifferenz reicht im Betrag von 8,37 bis 17,3. Das weitreichende HDI der Normalisierung weist indes darauf hin, dass die resultierenden *Posteriori*-Verteilungen in der Flachheit ihrer Ränder relativ unbestimmt bleiben. Insgesamt neigt die Analyse der MCMC-Approximation und die der grafisch manifesten *Posteriori*-Verteilung zwar zu einer Bestätigung der ursprünglichen Ergebnisse von Dauer et al. (2006), allerdings zu einer eher vorsichtigen Interpretation der tatsächlichen Effektstärke und daher des Ausmaßes der Interventionswirksamkeit, obwohl die *priors* eher nicht-informativen Verteilungen, insbesondere für die Standardabweichungen, entsprechen. Vor dem Hintergrund des angenommenen *priors* wird noch ein weiteres Erkenntnis gewonnen: Das Ergebnis ist in keinsten Weise überraschend. Durch den klassischen T-Test und die Annahme einer Nullhypothese (kein Effekt) wird mitunter ausgeblendet, dass ein positivistisches Studiendesign vorliegt. Dieses ist mindestens seit Karl Popper in seinem Maß an Erkenntnisgewinn fraglich. Der bayesianische Ansatz bietet also gleichzeitig eine Folie zur kritischen Selbstreflektion. Eine besondere Herausforderung für weitere Berechnungen, die den Umfang dieser Arbeit übersteigen, könnten indes die möglichst exakte mathematische Bestimmung vorhandener *priors* sein, die mitunter eine große Bedeutung für die Art bayesianisch kalkulierter Ergebnisse besitzen.

Die Exploration unterschiedlicher bayesianischer Verfahren macht deutlich, dass die grundlegende bayesianische Idee mit Hilfe des Satzes von der bedingten Wahrscheinlichkeit realisiert werden kann. Eindeutige Äquivalente zu bereits etablierten und verbreiteten Verfahren wie bspw. dem T-Test oder der Linearen Regression können jedoch auf Grund eines

inexistenten Konsenses nicht definiert werden. Vielmehr scheint die Ausdifferenzierung unterschiedlicher Verfahren auch in unterschiedliche Richtungen mit unterschiedlichen Schwerpunkten zu verlaufen (wie bspw. gut an Hand von den durch Gronau et al. (2017) aufgezeigten Diskursen über den Bayes-Faktor im Kontext von T-Tests oder dem aufwändigen Werk von Phadia (2016) über die komplexe Selektion von *priors* illustriert werden kann). Dies ist sicherlich partiell auf die relativ kleine Menge von Beiträgen in der Sozialwissenschaft, insbesondere in der Pflegewissenschaft zurückzuführen. Eine relevante Schlussfolgerung aus der Darstellung unterschiedlicher Verfahren besteht also aus dem transzendenten hohen Anspruch beim Gebrauch von Verfahren, die sich auf die Idee von Bayes berufen. Der kritisch-reflektierte Gebrauch dieser Verfahren mag dem historisch geformten Pragmatismus der Statistik in einigen Fällen entgegenstehen, enthält allerdings wie von Kruschke (2013) indirekt erwähnt eine gewisse Notwendigkeit, da ein standardisiertes Vorgehen dem Geist der bayesianischen Idee zuwiderläuft, bzw. der Philosophie subjektiver Transparenz sogar widerspricht. Bei der Anwendung bayesianischer Verfahren im Rahmen einer Sekundärdatenanalyse wurde ersichtlich, dass komplexere bayesianische Modelle im Vergleich klassisch frequentistischer eher zur Zurückhaltung neigen. Durch die Darstellung von Ergebnisverteilungen sind eindeutige Interpretationen an ein besonders hohes Maß verlässlicher Effekte gebunden. Dabei ist zu erkennen, dass die Interpretation skalarer Modelle (Bayes-Faktor) unter Umständen ähnliche Sensibilität für große Stichproben besitzt wie klassisch frequentistische. Gerade die mathematische Bestimmtheit von *priors* und die Offenlegung subjektiven Vorwissens stellen anscheinend für eine umfassendere Datenanalyse wichtige Voraussetzungen dar. Für eine differenzierte Interpretation von bayesianisch entwickelten Ergebnissen ist indes ein nicht triviales Wissen mathematischer Prozesse notwendig, wie gezeigt wurde. Grundvoraussetzung für einen vermehrten Einsatz solcher Verfahren ist aber eine wahrgenommene Relevanz dieser in disziplinbezogenen Fragestellungen. Im letzten Abschnitt dieser Arbeit soll daher die Frage nach der Relevanz oder dem Potential bayesianischer Statistik für die empirische Pflegeforschung beantwortet werden. Dies ist nun, nach einer

teilweise sehr formalen Darstellung der Differenz zur klassisch frequentistischen Statistik erst umfänglich möglich.

6. Disziplinbezogene Verortung

In diesem Teil der Arbeit soll das Potential bayesianischer Statistik für die empirische Pflegeforschung genauer untersucht werden. Dazu werden zuerst allgemeinere, erkenntnistheoretische Aspekte auf den Ansatz von Thomas Bayes bezogen. Im Anschluss wird das Verhältnis von Pflegewissenschaft und Statistik allgemein diskutiert, bevor auf konkretere theoretische Modelle, wie etwa das der pflegerischen Professionalisierung eingegangen wird.

6.1 Erkenntnis und die Definition von Wahrscheinlichkeit

In der historischen Betrachtung dieser Arbeit konnte gezeigt werden, dass die Suche nach einem idealen, allumfassenden Objektivismus ein Grund für den Rückzug einer subjektivistisch orientierten Statistik und einen ebenso subjektivistischen Wahrscheinlichkeitsbegriff war. Der Wahrscheinlichkeitsbegriff einer bestimmten Philosophie stellt sich letztendlich dar als ein Erklärungsmodell für das Auftreten, bzw. das Vorhersagen von Ereignissen einerseits und der damit verbundenen Frage nach legitimem, normativ-methodisch erlangtem Wissen. In der Pflegewissenschaft findet sich auch in der qualitativen Forschung eine Diskussion um die Art und Weise, wie Wissen am besten und sichersten, am wissenschaftlichsten generiert werden kann. Ein besonderes Augenmerk ist auch hierbei auf die Explikation des vorhandenen Wissens am Anfang einer Forschungsabsicht oder die Involvierung des Forschers in Form seines subjektivistischen Selbsts gerichtet. Die Phänomenologie und darauf aufbauend die Leibphänomenologie besitzt einen nicht unwesentlichen Einfluss auf die Methodologie der Pflegewissenschaft (vgl. Uzarewicz & Uzarewicz, 2005). Motiviert durch ihr Husserl'sches Erbe

thematisiert auch sie den Einfluss des eigenen Wissens und der im Forscher manifesten Interpretationsstrukturen auf das Ergebnis, das am Ende des Forschungsprozesses steht: Husserl (2009) fordert zur Annäherung an den Gegenstand des Interesses vom Forschenden eine Atheorie, das Beiseitelegen jeglichen bekannten und mutmaßlichen Wissens (S. 15 ff.). Die Negation jeglicher subjektiver Vorinterpretation und die Entwicklung der Statistik zu einer allgemeinen Erkenntnismethode ohne notwendigen Bezug zum Explorandum können vor dem Hintergrund einer Suche nach dem möglichst Objektiven interpretiert werden. Auch in der heutigen qualitativen Forschung ist das Prinzip oder der Anspruch an eine vollkommene Distanz zum Forschungsgegenstand noch präsent: Breidenstein, Hirschhauer, Kalthoff & Nieswand (2013) formulieren im Kontext ethnografischer Forschung den Anspruch an eine „*intellektuelle Distanz* zu der unmittelbaren Erfahrung des Feldes“ (S.109). Auch in der Sequenzanalyse wird bspw. von Maiwald gefordert, dass subjektives Wissen nicht in den primären Interpretationsprozess einfließen sollte (vgl. 2005). Der Wunsch nach Unvoreingenommenheit und einem Ideal der reinen Rationalität ist prinzipiell also in der heutigen Forschung weder im qualitativen noch im quantitativen Paradigma abwesend. Letztendlich bedingt die Transparenz der eigenen Vorannahmen jedoch die Interpretation anderer Forscher und der Wissenschaftsgemeinschaft. Der Vergleich der eigenen subjektiven Annahme und der empirisch belegten Erkenntnis garantiert erst eine Einschätzung über *hinzugewonnene Erkenntnis*, einer post-hoc Bewertung der Sinnhaftigkeit dieser und weiterer so gearteter Forschung. In diesem Punkt scheint die subjektive Transparenz aber eher mitunter als eine Schreckensgestalt zu wirken, die die Angst vor der Angreifbarkeit der eigenen Überzeugung mit sich führt.

In der Anwendung klassisch frequentistischer Verfahren zum Ausdruck eines ideellen Objektivismus kann dabei jedoch nicht von einer so ausgiebigen theoretischen Diskussion wie in der qualitativen Forschung gesprochen werden. Die zu Beginn dieser Arbeit erwähnte Selbsteinschätzung als einer reinen kritiklosen Methodologie haftet der frequentistischen Statistik immer noch an. Dabei mag jene gleichermaßen als Schutz vor externer Kritik und als Hindernis kritischer Selbstreflexion dienen. Für die Pflegewissenschaft bietet die Orientierung an

frequentistischen Verfahren zwar die Möglichkeit, eine gewisse wissenschaftliche und methodische Ausdifferenzierung im Bereich quantitativer Forschung zu demonstrieren, jedoch auch die Gefahr, ungeeignete Verfahren unreflektiert zu adaptieren. Gerade im Kanon einer *good clinical practice* oder einem *evidence based nursing* werden für den Goldstandard gehaltene randomisierte Kontrollstudien oft mit klassischen statistischen Verfahren, insbesondere dem T-Test, analysiert. Vor dem Hintergrund komplexer, multikausaler Interventionen, wie sie in der Pflege üblich sind, mag durch einen bayesianischen Ansatz jedoch eine differenziertere Erkenntnis über ein Phänomen möglich sein.

6.2 Empirische Pflegeforschung und Statistik

Durch die Literaturrecherche konnte keine deutsche Quelle für bayesianische Statistik identifiziert werden. Lediglich zwei deutsche Monographien wurden durch intensive Handsuche aufgefunden, die eher einen allgemeinen, nicht ausdifferenzierten und auch keinen praktischen Ansatz verfolgen. Erschöpft sich darin auch die Legitimität der vorliegenden Arbeit, so ist die Feststellung einer defizitären Praxis bayesianischer Statistik in Deutschland nicht unkritisch zu sehen. Lee & Abbott sprechen bereits 2003 in Kontexten der Pflegeforschung von einem komplexen Phänomenbereich, in dem bayesianische, multidimensionale Ansätze, insbesondere vor dem Hintergrund großer Stichprobenumfänge, angemessen erscheinen (vgl. S. 389). Die hierzulande verzögerte und verspätete Akademisierung der Pflege scheint sich teils in methodischer Zurückhaltung und Adaption aus angrenzenden Disziplinen zu manifestieren (vgl. Müller, 2011). Der durch bayesianische Statistik integrierbare Anteil theoretischen und/oder subjektiven Wissens mag im Rahmen komplexer (Sinn-)Konstruktionen wie der der Pflege eine besondere Bedeutung gewinnen, da eine verteilungsbedingte Analyse bestimmter Variablen und der Vergleich zu a priori definiertem Wissen mehr inhaltliche Aussagekraft für einen Phänomenbereich aufweist als bspw. das Testen atheoretischer (Null-)Hypothesen. Letztendlich mag darauf gehofft werden, dass die methodische Differenzierung im Feld der

quantitativen Pflegeforschung in Deutschland so weit voran schreitet, dass eine ausgiebige Diskussion über unterschiedliche statistische Schulen realistisch wird. Aus Sicht des Autors ist hierbei gerade eine universitäre Verortung der Pflegewissenschaft essentiell.

6.3 Professionalisierung der Pflege

Als klassische wissenschaftliche Explikation von Vorwissen werden normalerweise ebenfalls wissenschaftliche Veröffentlichungen und/oder Ergebnisse aus Datenerhebungen genutzt, sofern sie für ein bestimmtes Thema oder Phänomen existieren. Selbst die Expertenmeinung des Forschers fußt mutmaßlich auf methodisch begründeter Empirie und der Analyse wissenschaftlicher Literatur. Sollte diese Vorgehensweise zur Definition eines *priors* für eine bayesianische Statistik auch einem gewissen Standard entsprechen (vgl. bspw. Sawka et al., 2007, S. 338 f.), besagt der *Satz von der bedingten Wahrscheinlichkeit* nicht, woher das Wissen eines *priors* stammt. Nimmt man Thomas Bayes in seiner Absicht ernst, einen Glaubensgrad mit in die Interpretation eines Phänomens einzubeziehen, läuft dies nicht zwangsläufig auf eine wissenschaftliche Legitimität hinaus. Wie in vorherigen Kapiteln gezeigt wurde, war und ist dies der Grund für die wissenschaftliche Anfechtung bayesianischer Methodik. Während Vertreter bayesianischer Statistik ihr Vorgehen mit bereits etablierten wissenschaftlichen Praktiken begründen, indem sie a priori Wissen aus anerkannter Literatur und Datenauswertungen exzerpieren, drängt sich die Frage nach dem Verhältnis wissenschaftlichen Wissens und Erfahrungswissens in dem Kontext bayesianischer Analyse geradezu auf. Die Möglichkeit, Erfahrungswissen als *prior* zu definieren steht in keinem direkten Widerspruch zu der Philosophie, die sich hinter dem Ansatz bayesianischer Statistik verbirgt. Auch mathematisch-operativ stellt sich diese Frage vor dem Hintergrund rechenintensiver, aber machbarer Approximations- und Regressionsverfahren zur Simulation mathematisch einwandfreier Verteilungen nicht. Auch hier hilft ein Blick auf gängige qualitative Forschung, um zu demonstrieren, dass die empirische Offenlegung subjektiven, phänomenbezogenen, aber wissenschaftlich

intransparenten Wissens nicht der Verlust wissenschaftlicher Anerkennung bedeuten muss. Qualitative Experteninterviews können hierfür exemplarisch angeführt werden: Hier besteht die Annahme, dass auch das Wissen von Personen, die keine wissenschaftliche Ausbildung absolviert haben, zu dem Erkenntnisgewinn über einen Gegenstand beiträgt, sofern die *Analyse* und *Verarbeitung* dieser Informationen mit Hilfe wissenschaftlicher Methodik geschieht und die Befragten tatsächlich der Anforderung eines Experten genügen (vgl. bspw. Kaiser, 2014, S.35 ff.). Im Kontext bayesianischer Schätzungen bedeutet dies unter anderem, nicht-wissenschaftliches, subjektives Vorwissen in eine verfahrensbezogen angemessene und mathematisch widerspruchsfreie Form zu überführen, sich aber nicht der Scheu vor oder der Hybris über eben diesem nicht-wissenschaftlichen Wissen zu ergeben. Außerdem muss klar sein, dass die Quelle des subjektiven Wissens der bereits beschriebenen Anforderung genügt, sich tatsächlich angemessen über ein Phänomen äußern zu können (außer die Forschungsfrage richtet sich konkret auf Gegenteiliges, etwa Laienwissen).

Das Potential bayesianischer Statistik liegt in ihrem analytisch-logischen Vergleich von Vorwissen und Wissen, welcher mit Hilfe von erhobenen Daten interpretiert wird. Pflegekräfte nach ihrer Erfahrung über ein bestimmtes Phänomen zu fragen und die Ergebnisse daraus bspw. mit literaturgestützten Vorannahmen oder normativ vorgegebenen Annahmen abwägen zu können, stellt eine innovative Form der statistischen Analyse dar. Es passt darüber hinaus sehr gut mit dem aus der Pflegewissenschaft formulierten Anspruch an professionelles Handeln zusammen. Weidner (1995) hat ausführlich die doppelte Handlungslogik Ulrich Oevermanns dargelegt und auf die pflegerische Praxis übertragen (vgl. S. 48 ff.). Professionelles Handeln stellt sich folglich als ein zeitweise unter Handlungszwang stehendes Abwägen zwischen fallbezogenem, individuellem und subjektivem Wissens sowie einer theoretischen Normation dar. Bayesianische Statistik ist in der Lage, subjektives Handlungswissen einerseits und vorgegebenes, theoretisches Wissen andererseits zu relationieren. Die Verlässlichkeit theoretischer Implikationen vor dem Hintergrund bestimmter Phänomene zu analysieren, stellt im Grunde ein bayesianisches Hypothesentesten dar. Durch den

Vergleich verschiedener Ergebnisse unterschiedlicher *priors*, die mit den gleichen Daten kombiniert werden, könnten bspw. auch Unterschiede in Kompetenzen zur Vorhersage von Krankheitsverläufen oder ähnlichem aufgezeigt werden. Dadurch könnte sogar eine empirisch differenzierbare Abbildung differenter Professionalisierungsniveaus oder des Verständnisses von Pflege an sich analysiert werden. Gerade in Bezug auf einen handlungsorientierten Professionsansatz der Pflege gilt es jedoch, einen überdosierten Pragmatismus statistischer Erkenntnisproduktion zu vermeiden und die Defizite dieses Ansatzes nicht wissenschaftlich zu legitimieren. So spannend die Relation praktischen und theoretischen Wissens auch sein mag, sollte in einer so gearteten Forschung berücksichtigt werden, dass Praxis auch immer strukturell bedingt ist und nicht zwangsläufig durch die persönliche Überzeugung des Handelnden impliziert wird. Durch kritische Professionstheorien, die etablierte Strukturen und Machtverhältnisse reflektieren (bspw. Kellner, 2011), kann Handeln selbst in Relation gestellt werden. Gerade die zur Selbstreflektion anregende bayesianische Statistik sollte deshalb das Setting, in der sie angewendet wird, diskutieren. Gerade in der bayesianischen Analyse von Pflegepraxis, die nach Ansicht des Autors durch teilweise theoretisch unbegründete normative Direktiven geprägt ist, sollte dieser Aspekt berücksichtigt werden. Ansätze für erste Analysen durch bayesianische Statistik sind derweil in der empirischen Pflegeforschung vorhanden. Ein bayesianisch strukturierter Vergleich von durch Experten erklärte Pflegezeit mit theoretisch oder normativ erklärter Pflegezeit könnte an bereits etablierte Studien und Forschungsrichtungen (z.B. Brühl & Planer, 2013) anschließen. Auch die Vorhersage von Krankheitsverläufen und die Entwicklung von Krankheitsmodellen erscheint ein vielversprechender Ansatz, sofern statistisch komplexe Modelle, wie bspw. *Recurrent Neural Networks* für Krankheitsverläufe (vgl. Nguyen, Tran & Venkatesh, 2017) mit bayesianischen Netzwerken (vgl. Lee & Abott, 2003) kombiniert werden können.

7. Diskussion und Limitation

Eine Analyse der gängigen Literatur hat gezeigt, dass die bayesianische Statistik in der empirischen Pflegeforschung erst in der Entwicklung begriffen ist, insbesondere in Deutschland. Die ohnehin methodisch vielfältig orientierte Disziplin Pflegewissenschaft findet in der bayesianischen Statistik zu allererst einen weiteren Anreiz zur methodischen Differenziertheit, aber auch zur Zerstreuung. Die Frage, die sich stellt, ist unter anderem: Soll eine Disziplin eher gemäßigter statistischer Euphorie methodologisch so stark irritiert werden, dass man sie vor die Wahl „entweder frequentistisch oder bayesianisch“ stellt? Aus Sicht des Autors geht die inhaltliche Tiefe bayesianischer Statistik über die eines rein empirisch-analytischen Paradigmas hinaus. Sie wirft viel mehr grundsätzliche Fragen über die Struktur empirischer Forschung auf, Fragen, die sich auf die Voreingenommenheit und das Vorwissen des Forschers richten und sich letztendlich zu der Aufforderung zur Transparenz entwickeln.

Mathematisch-operationell ist bayesianische Statistik von zwei Seiten zu betrachten. Ihre Anwendung erfordert das umfängliche Einlesen in einen eher spezialisierten Bereich, der von der Wissenschaftsgemeinschaft nicht so weit diskutiert wurde, dass eine klare Bewertung seines Potentials möglich ist. Darüber hinaus sind die so erlangten Ergebnisse meist weniger eindeutig als diejenigen frequentistischer Verfahren. Die Darstellung linearer Regression aus bayesianischer Sicht hat bspw. gezeigt, dass es nicht um die Festlegung bestimmter Werte für Regressionskoeffizienten oder -gewichte geht, sondern um eine funktionsanalytische Interpretation von bestimmten Intervallen von Regressionskoeffizienten, also interpretierbaren, mathematischen *Spielräumen*. In der Konsequenz muss sich der Forschende von solch eindeutigen Verfahren wie Signifikanztests distanzieren. In der handlungsorientierten oder sogar unter Handlungszwang stehenden Disziplin der Pflege (oder Medizin) kann eine gewisse Eindeutigkeit jedoch eine bestimmte Attraktivität besitzen. Dabei ist die mathematische Ungenauigkeit von einer anderen Seite betrachtet als ein Anreiz spezifischerer und aufeinander aufbauender Forschung

interpretierbar. Hinsichtlich eines Prozesses iterativen Lernens ist sicherlich, ebenso wie in der einmaligen Anwendung bayesianischer Verfahren, die Gewichtung zwischen dem *prior* und den gegebenen Daten zu diskutieren. Die Neigung bestimmter Verfahren, bei großen Stichproben *prior*-unabhängige Ergebnisse zu erzielen, ist kritisch zu betrachten. Nichtsdestotrotz ist aber deutlich, dass der Anspruch an statistische Analyse durch eine Stetigkeit von Interpretation und nicht durch ihre Dichotomie (im Sinne signifikant/nicht signifikant) gehoben werden kann. Dadurch ist zwar tendenziell der Umgang mit höherer Komplexität erforderlich, der oftmals erwähnten Informations- und Datenflut der (Pflege-)Wissenschaft mag dies allerdings zu mehr Ordnung verhelfen.

Ideell-philosophisch kann der bayesianische Ansatz ebenfalls diskutiert werden. Thomas Bayes ging davon aus, dass die Beurteilung der Wahrscheinlichkeit eines Ereignisses von dem Glaubensgrad (oder dem subjektiven Vorwissen) des Beurteilenden abhängig ist. In post-sozialkonstruktivistischen Zeiten ist dieser Ansatz sehr gut nachvollziehbar. Erhobene Daten vor dem Hintergrund des Vorwissens (die Definition der *Likelihood*) in eine statistische Analyse zu integrieren, mag jedoch nicht direkt mit einer bereits erwähnten phänomenologischen Überzeugung überein gehen. Aus Sicht des Autors ist in diesem Kontext die Differenzierung von subjektiver Transparenz und subjektiver Erwartung entscheidend. Subjektive Transparenz besteht dabei eher in der Offenlegung und möglichst detaillierten (mathematischen) Beschreibung der subjektiven Überzeugung über ein Phänomen, subjektive Erwartung eher in einer deterministischen Beeinflussung von Ergebnissen eines Erkenntnisprozesses. Auch, wenn bayesianische Verfahren die Definition von *priors* fordern, so ist erstens zu sehen, dass einige Wege darum herum führen (nicht informative *priors*) und zweitens erst die Interpretation des Ergebnisses (*Posteriori*-Verteilung) subjektiv-deterministische Züge aufweisen kann. Die Berechnung einer *Posteriori*-Verteilung und die Bestimmung eines 95%-HDIs, ja sogar die Kalkulation eines Bayes-Faktors allein stellt kein Kriterium für die Relevanz oder Irrelevanz des Ergebnisses dar, es ist in erster Linie eine formal logische und klar methodisch definierte Kalkulation. Die Zuweisung von normativen Definitionen eines bestimmten Relativwertes stellt erst den Zwang wissenschaftssoziologischer

Mechanismen her. Diese führen dann u.U. zu solchen Ausprägungen wie der Litanei von 5%-Signifikanzgrenzen, können aber genauso gut zu einer „Bayes-Faktor: 20“- Grenze werden. Wichtig ist also, bayesianischer Statistik keine Unwissenschaftlichkeit per se durch die Formulierung subjektiven Wissens zu attestieren, weder aus klassisch frequentistischer, noch aus qualitativ forschender Perspektive. Viel eher kann sie zu einer methodischen Diversität beitragen und die Frage nach setting-abhängiger Angemessenheit bestimmter statistischer Methoden stellen, indem sie sich ein Stück weit von der heutigen Definition eines analytischen Empirismus (wie ihn bspw. Friesacher, 2008 beschreibt) entfernt.

Die vorliegende Arbeit kann in ihrem Umfang diese höchst spannenden philosophischen und mathematischen Diskussionen nicht umfänglich aufgreifen. Sie ist als Einführung in den Themenbereich bayesianischer Statistik gedacht und weist daher Limitationen auf. Die historische Betrachtung am Anfang dieser Arbeit wurde bewusst kurz gehalten und möglichst komprimiert dargestellt. Dadurch gingen zwangsläufig Details verloren. Der mathematisch-operationale Teil hielt lediglich eine Auswahl der gängigsten Verfahren bayesianischer Statistik bereit und eine erste vorsichtige exemplarische Berechnung. Auch in dieser Hinsicht ist eine weitere Ausarbeitung komplexerer Verfahren denkbar. Zu guter Letzt stellt sich die disziplinbezogene Verortung als eine Reihe von Impulsen dar, die dem Autor wichtig erschienen. Jene ist eher als der Wunsch nach einer anschließenden Diskussion und einer gemeinsamen Reflektion zu interpretieren als die Abgrenzung bestimmter Themen oder das Einfordern erkenntnistheoretischer Monopole. Diese Arbeit kommt somit dem in der Literatur exploriertem Defizit nach, grundlegende Ideen und einführende Verfahren bayesianischer Statistik vorzustellen. Im letzten Kapitel dieser Arbeit sollen eine kurze Zusammenfassung erfolgen und Perspektiven für anschließende Arbeiten eröffnet werden.

8. Zusammenfassung und Ausblick

Nur wenige Arbeiten zum Thema der Anwendung bayesianischer Verfahren in der Pflegewissenschaft waren zu finden. Methodische Differenziertheit wiesen eher monographische Werke auf, die an Sozialwissenschaftler im Allgemeinen adressiert waren. Die grundlegende Idee bayesianischer Statistik wurde vor dem historischen Kontext als eine zu der entsprechenden Zeit nicht unübliche Widerspruchsfreiheit aus subjektiver und objektiver Erkenntnis dargestellt. Erst durch den Übertrag subjektiven Verstehens auf die objektivierende Disziplin der Mathematik und die darauf folgende Re-Duktion eines objektiven Anspruchs an empirische Methodik bedingte die aufkommende Kritik an subjektivistischer Statistik im philosophischen Umfeld eines radikalen Rationalismus. Obwohl die bayesianische Idee im Verlauf der Geschichte kurze Momente der Rekapitulation hatte, entwickelte die Statistik eher Züge einer illusionsreichen Omnipotenz durch rationale Logik und mathematische Unanfechtbarkeit. Erst ab den 1970er Jahren wurde das Potential bayesianischer Schätzung vor dem Hintergrund komplexerer Analysen und des informationstechnischen Fortschritts wieder klar in die Wissenschaftsdiskussion integriert. Stellt sich der Satz der bedingten Wahrscheinlichkeit auf den ersten Blick als simpel dar, erweist sich eine tatsächliche Kalkulation von mehrdimensionalen, stetigen Phänomenen als nicht trivial (was wohl der Realität wesentlich näher kommt als eindimensionale, diskrete Vorstellungen von ihr). MCMC-Verfahren, im Wesentlichen Algorithmus gestützte, iterative Approximationsverfahren, sind zur gängigen Lösung dieses Problems geworden. Durch die Beschreibung einiger wichtiger Verfahren und der exemplarischen Anwendung zweier dieser Verfahren sollte dem Leser ein erster Eindruck einer alternativen, verteilungsbasierten Analyse vermittelt werden, die weniger eindeutig erscheint als manch frequentistisches Pendant, dafür umso angemessener erscheint für den Zugewinn an Erkenntnis über hierarchische, mehrdimensionale und komplexere Phänomene. In den Grundzügen konnte die Bedeutung bayesianischer Statistik und ihr Potential für die empirische Pflegeforschung in Form einiger Aspekte illustriert werden. Die Einschränkungen der vorliegenden Arbeit und das

Für und Wider ihrer hypothetischen Tragweite wurden abschließend offen gelegt.

Perspektiven für anknüpfende Arbeiten sind zahlreich. Im Sinne eines mathematisch-operationalen Ausblicks ließen sich auch nicht-parametrische bayesianische Verfahren genauer explorieren. Im Abschnitt „Bayesianische Mehrebenenanalyse“ wurden Annahmen über bestimmte Verteilungen zur Berechnung von *Posteriori*-Verteilungen getroffen (multinomiale Verteilung, dem i.i.d.-Kriterium unterliegend). Diese Annahmen sind ähnlich wie in der frequentistischen nicht-parametrischen Statistik jedoch nicht zwangsläufig angemessen oder nachvollziehbar. Bayesianische nicht-parametrische Modelle halten Ansätze bereit, die Verteilungen für Verteilungen simulieren (bspw. mit Hilfe des sog. *Dirichlet-Prozesses*, vgl. Müller, Quintana, Jara & Hanson, 2015). Wie im Kapitel 5.3 „Die Markow Chain Monte Carlo Methode“ bereits erwähnt, stellt außerdem die Exploration unterschiedlicher Verfahren zur Bestimmung von *Posteriori*-Verteilungen eine spannende Diskussion innerhalb bayesianischer Statistik dar. Nicht nur die genauere Aufzeichnung mehrerer verschiedener Algorithmen für das MCMC-Verfahren, sondern auch Alternativen für jene versprechen die Darstellung eines umfangreichen, fast schon separaten Themenbereiches mit einer starken mathematischen Ausrichtung, inkl. numerischer Integration durch verschiedene Interpolationsarten (vgl. Richter & Wick, 2017, S. 385 ff.). Dazu kommen Themen wie die unterschiedlichen Selektionen und Bestimmungen von *priors* und eine damit verbundene kritische Reflektion pragmatischer Statistik oder Darstellung unterschiedlicher computerbasierter Programme und ihrer Vor- und Nachteile, wie bspw. R (und dessen unterschiedliche packages), SPSS, WinBUGS oder STATA, sowie wenig bekannte Programme, die im Zuge bayesianischer Statistik erst entstanden, wie bspw. LePAC.

Einzelne Konnotationen und Strömungen innerhalb der bayesianischen Statistik aufzuzeigen, stellt einen weiteren Anreiz differenzierterer und umfänglicherer Arbeiten dar (die den Umfang einer Einführung überstiegen hätte). Auch die bayesianische Statistik unterliegt einem anhaltenden Diskurs, in dem vor allem in subjektive und objektive bayesianische Statistik unterteilt wird. Was vor dem historischen Kontext als widersinnig

erscheinen mag, erhält seine Relevanz bspw. in der Wahl von informativen oder nicht-informativen *priors* und hat daher eine gewisse Auswirkung auf die Strukturierung und Prozessierung bestimmter Verfahren. Selbst in den einzelnen Unterkategorien des bayesianischen Diskurses sind wiederum unterschiedliche Ansichten über grundständige Annahmen zu finden (z.B. die Diskussion über die Interpretation und Berechnung der *Likelihood*, wie sie die *fiducial bayesian inference* aufwirft, vgl. Lecoutre & Pointevineau, 2014, passim). Eine Systematisierung und Ordnung dieser Strömungen mag in Form eines umfänglichen scoping reviews zur Klärung einer für die empirische Pflegeforschung besser oder weniger geeigneten Orientierung innerhalb bayesianischer Verfahren dienen. Auch eine kritische und selbstreflektierende Aufarbeitung der Historie der Statistik stellt eine enorm anspruchsvolle, aber aus Sicht des Autors wichtige Perspektive dar. Die Möglichkeiten zur Erweiterung der vorliegenden Arbeit scheinen vielfältig. Letztendlich ist auch die empirische Anwendung unterschiedlicher Verfahren essentiell für die praktische und pragmatische Beurteilung.

Bayesianische Statistik soll in ihrem Wesen im Sinne des Titels dieser Arbeit interpretiert werden. Es handelt sich um eine methodische Alternative zu elaborierten statistischen Verfahren. Sie erhebt keinen Anspruch darauf, *erhaben* zu sein über all die Methoden, die durch beeindruckende menschliche Denkprozesse zur Erweiterung unserer Erkenntnis beigetragen haben und immer noch beitragen. Nach Meinung des Autors bestehen für die Pflegewissenschaft eine gewisse Relevanz und das aufgezeigte Potential in der durch Subjektivität bedingten Kontrastierung wissenschaftlicher Forschung. Es handelt sich dabei nicht um die von Zeit zu Zeit immer wieder aufkommende Erklärung über die Auflösung oder Synthese qualitativer und quantitativer Forschung oder des analytischen und phänomenologisch-hermeneutischen Paradigmas, sondern um die Weiterentwicklung und Ausdifferenzierung empirischer Forschung, ohne sie in eine Chimära omnipotenter stochastischer Methodik verwandeln zu wollen. Zufall ist laut Albert Einstein schließlich das, wobei unsere Berechnungen versagen.

Literaturverzeichnis

- Albert, J. (2018). LearnBayes: Functions for Learning Bayesian Inference: R package version 2.15.1. Retrieved from <https://CRAN.R-project.org/package=LearnBayes>.
- Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K., & Stachel, H. (2015). *Mathematik* (3. Auflage). Berlin, Heidelberg: Springer Spektrum.
- Bättig, D. (2017). *Angewandte Datenanalyse: Der Bayes'sche Weg* (2., überarbeitete und erweiterte Auflage). *Statistik und ihre Anwendungen*. Berlin: Springer Spektrum.
- Berlowitz, D. R., Christiansen, C. L., Brandeis, G. H., Ash, A. S., Kader, B., Morris, J. N., & Moskowitz, M. A. (2002). Profiling nursing homes using Bayesian hierarchical modeling. *Journal of the American Geriatrics Society*, 50(6), 1126–1130.
- Bortz, J., & Schuster, C. (2016). *Statistik für Human- und Sozialwissenschaftler: Extras online* (Limitierte Sonderausgabe, 7., vollständig überarbeitete und erweiterte Auflage). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer.
- Breidenstein, G., Hirschauer, S., Kalthoff, H., & Nieswand, B. (2013). *Ethnografie: Die Praxis der Feldforschung. UTB Sozialwissenschaften, Kulturwissenschaften: Vol. 3979*. Konstanz: UVK Verl.-Ges.
- Brokate, M., Henze, N., Hettlich, F., Meister, A., Schranz-Kirlinger, G., & Sonar, T. (2016). *Grundwissen Mathematikstudium: Höhere Analysis, Numerik und Stochastik*. Berlin: Springer.
- Brühl, A., & Planer, K. (2013). PiSaar: Pflegebedarf im Saarland - Abschlussbericht. Retrieved from http://opus.bsz-bw.de/kidoks/volltexte/2013/117/pdf/PiSaar_Abschlussbericht_2013.pdf.
- Büchter, A., & Henn, H.-W. (2007). *Elementare Stochastik: Eine Einführung in die Mathematik der Daten und des Zufalls* (Zweite, überarbeitete und erweiterte Auflage). *Mathematik für das Lehramt*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Cowles, M. K. (2013). *Applied Bayesian Statistics: With R and OpenBUGS Examples. Springer texts in statistics: Vol. 98*. New York, NY: Springer.
- Dauer, L. T., Kelvin, J. F., Horan, C. L., & St Germain, J. (2006). Evaluating the effectiveness of a radiation safety training intervention for oncology nurses: a pretest-intervention-posttest study. *BMC Medical Education*, 6(32), 1-10.
- Friesacher, Heiner (2008). *Theorie und Praxis pflegerischen Handelns: Begründung und Entwurf einer kritischen Theorie der Pflegewissenschaft*. Göttingen: V & R Unipress.
- Gentle, J. E., Härdle, W. K., & Mori, Y. (Eds.). (2012). *Handbook of Computational Statistics: Concepts and Methods* (2nd ed. 2012. revised and updated.). *Springer Handbooks of Computational Statistics*. Berlin, Heidelberg: Springer.
- Gigerenzer, G. (1989). *The empire of chance: How probability changed science and everyday life. Ideas in context: Vol. 12*. Cambridge: Cambridge University Press.
- Gönen, M., Johnson, W. O., Lu, Y., & Westfall, P. H. (2005). The Bayesian Two-Sample t Test. *The American Statistician*. (59), 252–257.
- Goodson, J., Jang, W., & Rantz, M. (2008). Nursing home care quality: insights from a Bayesian network approach. *The Gerontologist*, 48(3), 338–348.

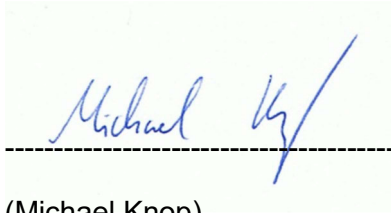
- Gronau, Q. F., Ly, A., & Wagenmakers, E.-J. (2017). Informed Bayesian T-Tests. Retrieved from <https://arxiv.org/abs/1704.02479>.
- Hanson, K. M., & Wolf, D. R. (1996). Estimators for the Cauchy Distribution. *Maximum Entropy and Bayesian Methods*. (07), 255–263. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.19.1568&rep=rep1&type=pdf>.
- Held, L., & Bové, D. S. (2014). *Applied statistical inference: Likelihood and Bayes*. Heidelberg u.a.: Springer. Retrieved from <http://d-nb.info/1032794003/04>.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer texts in statistics. Dordrecht u.a.: Springer.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2. ed.). Quantitative methodology series. New York, NY: Routledge.
- Husserl, E. (2009). *Allgemeine Einführung in die reine Phänomenologie*. Nachwort (1930): [text- und seitengleich nach der kritischen Edition in Husserliana III/1, Den Haag 1976 und Husserliana V, Den Haag 1971]. Philosophische Bibliothek: Vol. 602. Hamburg: Meiner.
- Jackman, S. (2009). *Bayesian analysis for the social sciences*. Wiley series in probability and statistics. Chichester: Wiley.
- Kaiser, R. (2014). *Qualitative Experteninterviews: Konzeptionelle Grundlagen und praktische Durchführung*. Lehrbuch. Wiesbaden: Springer VS.
- Kaplan, D. (2014). *Bayesian Statistics for the Social Sciences*. Methodology in the Social Sciences. New York: Guilford Publications.
- Kellner, A. (2011). *Von Selbstlosigkeit zur Selbstsorge: Eine Genealogie der Pflege*. Zugl.: Freiburg (Breisgau), Pädag. Hochsch., Diss., 2011. *Pflege und Gesundheit: Bd. 4*. Berlin: Lit.
- Kent, B., & McCormack, B. (2011). *Clinical Context for Evidence-based Nursing Practice*. Evidence Based Nursing. Hoboken: John Wiley & Sons.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology. General*, 142(2), 573–603.
- Kruschke, J. K., & Meredith, M. (2018). BEST: Bayesian Estimation Supersedes the t-Test. Retrieved from <https://CRAN.R-project.org/package=BEST>.
- Lecoutre, B., & Poitevineau, J. (2014). *The Significance Test Controversy Revisited: The Fiducial Bayesian Alternative*. SpringerBriefs in Statistics. Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg.
- Lee, J., Sung, W., & Choi, J.-H. (2015). Metamodel for Efficient Estimation of Capacity-Fade Uncertainty in Li-Ion Batteries for Electric Vehicles. *Energies*, 8(6), 5538–5554.
- Lee, S.-M., & Abbott, P. A. (2003). Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. *Journal of Biomedical Informatics*, 36(4-5), 389–399.
- Lopes, M. V. d. O., da Silva, V. M., & Araujo, T. L. de. (2012). A análise de diagnósticos de enfermagem sob uma perspectiva bayesiana. *Revista Da Escola De Enfermagem Da U S P*, 46(4), 994–1000.
- Los Campos, G. de, & Rodriguez, P. P. (2014). Bayesian Linear Regression. Retrieved from <https://CRAN.R-project.org/package=BLR>.
- Maiwald, K.-O. (2005). Competence and Praxis: Sequential Analysis in German Sociology. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0114-fqs0503310>.

- McGrayne, S. B. (2012). *The theory that would not die: How Bayes' rule cracked the Enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy* (Paperback ed.). New Haven Conn. u.a.: Yale Univ. Press.
- Meredith, M., & Kruschke, J. K. (2018). HDInterval: Highest (Posterior) Density Intervals. Retrieved from <https://CRAN.R-project.org/package=HDInterval>.
- Misiorek, A., & Weron, R. (2012). Heavy-Tailed Distributions in VaR Calculations. In J. E. Gentle, W. K. Härdle, & Y. Mori (Eds.), *Springer Handbooks of Computational Statistics. Handbook of Computational Statistics: Concepts and Methods* (2nd ed., pp. 1025–1060). Berlin, Heidelberg: Springer.
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, 16(4), 406–419.
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factor for Common Design. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>.
- Müller, M. (2011). *Statistik für die Pflege: Handbuch für Pflegeforschung und -wissenschaft* (1. Aufl.). s.l.: Verlag Hans Huber.
- Müller, P., Quintana, F. A., Jara, A., & Hanson, T. (2015). *Bayesian Nonparametric Data Analysis. Springer Series in Statistics*. Cham, s.l.: Springer International Publishing.
- Neher, M. (2018). *Anschauliche Höhere Mathematik für Ingenieure und Naturwissenschaftler 1: Lineare Algebra, eindimensionale Analysis: Ein graphisch orientierter Zugang. Lehrbuch*. Wiesbaden: Springer Vieweg.
- Nguyen, P., Tran, T., & Venkatesh, S. (2017). *Finding Algebraic Structure of Care in Time: A Deep Learning Approach*. Retrieved from NIPS 2017 website: <http://arxiv.org/pdf/1711.07980v1>.
- Peltokoski, J., Vehviläinen-Julkunen, K., Pitkäaho, T., Mikkonen, S., & Miettinen, M. (2015). The comprehensive health care orientation process indicators explain hospital organisation's attractiveness: a Bayesian analysis of newly hired nurse and physician survey data. *Journal of Nursing Management*, 23(7), 954–962.
- Phadia, E. G. (2016). *Prior Processes and Their Applications: Nonparametric Bayesian Estimation* (Second edition). *Springer Series in Statistics*. Cham: Springer.
- Pitkäaho, T., Partanen, P., Miettinen, M., & Vehviläinen-Julkunen, K. (2015). Non-linear relationships between nurse staffing and patients' length of stay in acute care units: Bayesian dependence modelling. *Journal of Advanced Nursing*, 71(2), 458–473.
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>.
- Remmers, H. (2011). Pflegewissenschaft als transdisziplinäres Konstrukt. Wissenschaftssystematische Überlegungen - Eine Einleitung. In H. Remmers (Ed.), *Pflegewissenschaft im interdisziplinären Dialog. Eine Forschungsbilanz*. (pp. 7–47). Osnabrück: V&R.
- Remmers, H. (Ed.). (2011). *Pflegewissenschaft im interdisziplinären Dialog. Eine Forschungsbilanz*. Osnabrück: V&R.
- Richter, T., & Wick, T. (2017). *Einführung in die Numerische Mathematik: Begriffe, Konzepte und zahlreiche Anwendungsbeispiele*. Berlin, Heidelberg: Springer Spektrum.

- Samaniego, F. J. (2010). *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. Springer Series in Statistics. New York, NY: Springer Science+Business Media LLC.
- Sawka, A. M., Boulos, P., Beattie, K., Papaioannou, A., Gafni, A., Cranney, A., . . . Thabane, L. (2007). Hip protectors decrease hip fracture risk in elderly nursing home residents: a Bayesian meta-analysis. *Journal of Clinical Epidemiology*, 60(4), 336–344.
- Schäfer, T. (2010). *Deskriptive und explorative Datenanalyse* (1. Aufl.). Wiesbaden: VS Verl. für Sozialwiss.
- Schäfer, T. (2011). *Statistik II: Inferenzstatistik* (1. Aufl.). Wiesbaden: VS-Verl.
- Schäfer, T. (2016). *Methodenlehre und Statistik: Einführung in Datenerhebung, deskriptive Statistik und Inferenzstatistik* (1. Aufl. 2016). *Basiswissen Psychologie*. Wiesbaden: Springer.
- Tervo-Heikkinen, T., Kiviniemi, V., Partanen, P., & Vehviläinen-Julkunen, K. (2009). Nurse staffing levels and nursing outcomes: a Bayesian analysis of Finnish-registered nurse survey data. *Journal of Nursing Management*, 17(8), 986–993.
- Thompson, D. R., & Martin, C. R. (2017). Bayes' theorem and its application to cardiovascular nursing. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, 1474515117712317.
- Tschirk, W. (2014). *Statistik: Klassisch oder Bayes: Zwei Wege im Vergleich*. Springer-Lehrbuch. Berlin: Springer Spektrum.
- Uzarewicz, C., & Uzarewicz, M. (2005). *Das Weite suchen: Einführung in eine phänomenologische Anthropologie für Pflege. Bildung – Soziale Arbeit – Gesundheit: v.07*. Berlin: De Gruyter Oldenbourg.
- Wang, M., & Liu, G. (2016). A Simple Two-Sample Bayesian t -Test for Hypothesis Testing. *The American Statistician*, 70(2), 195–201.
- Weidner, F. (1995). *Professionelle Pflegepraxis und Gesundheitsförderung: Eine empirische Untersuchung über Voraussetzungen und Perspektiven des beruflichen Handelns in der Krankenpflege*. Zugl.: Osnabrück, Univ., Diss., 1994. Mabuse-Verlag Wissenschaft: Vol. 22. Frankfurt am Main: Mabuse-Verl.
- Ziegler, M. (2017). *Induktive Statistik und soziologische Theorie* (Dissertation). Juventa Verlag.

Einverständniserklärung

Ich bin damit einverstanden, dass meine Masterarbeit online über den Hochschulschriftenserver KiDokS oder in der Bibliothek der Philosophisch-Theologischen Hochschule Vallendar veröffentlicht und damit der Fachleserschaft zugänglich gemacht wird.

A handwritten signature in blue ink, appearing to read "Michael Knop", is written on a light green rectangular background. A horizontal dashed line is positioned below the signature.

(Michael Knop)

Bielefeld, den 08.08.2018

Anhang: Rechercheprotokoll

Datum	Datenbank	Suchbegriff	Suchart	Anzahl der Treffer	Anzahl relevanter Treffer	Einschluss	Bemerkung
22.04.2018	Livivo	bayes*	Titel	21718			Weitere Eingrenzung
22.04.2018	Livivo	bayes* AND statistic*	Titel	1024			Weitere Eingrenzung
22.04.2018	Livivo	bayes* AND statistic* AND nurs*	Titel	0			
22.04.2018	Livivo	bayes* AND statistic* AND social	Titel	23	7	3	in Datei LIVIVO_1
22.04.2018	Livivo	bayes* AND histori*	Titel	62	1	0	
22.04.2018	Livivo	bayes* AND social AND science	Titel	42	14	10	in Datei LIVIVO_2
22.04.2018	Livivo	bayes* AND analysis	Titel	3622			Weitere Eingrenzung
22.04.2018	Livivo	bayes* AND analysis AND nurs*	Titel	4	3	3	in Datei LIVIVO_2
22.04.2018	Livivo	bayes* AND analysis AND social	Titel	20	0		
22.04.2018	Livivo	bayes* AND nurs*	Titel	20	13	9	in Datei LIVIVO_3
22.04.2018	Livivo	bayes* AND statistik	Titel	1165			Weitere Eingrenzung
22.04.2018	Livivo	bayes* AND statistik AND pfleg*	Titel	0			
22.04.2018	Livivo	bayes* AND statistik AND sozial	Titel	0			
22.04.2018	Livivo	bayes* AND statistik AND sozial*	Titel	0			
22.04.2018	Livivo	bayes* AND statistik AND Einführung	Titel	21	3	2	in Datei LIVIVO_2

Datum	Datenbank	Suchbegriff	Suchart	Anzahl der Treffer	Anzahl relevanter Treffer	Einschluss	Bemerkung
04.05.2018	PubMed	bayes*	Titel	11262			
04.05.2018	PubMed	bayes* AND statistic*	Titel	303			
04.05.2018	PubMed	bayes* AND statistic* AND nurs*	Titel	0			
04.05.2018	PubMed	bayes* AND statistic* AND social	Titel	0			
04.05.2018	PubMed	bayes* AND histori*	Titel	37	0		
04.05.2018	PubMed	bayes* AND social AND science	Titel	0			
04.05.2018	PubMed	bayes* AND analysis	Titel	2088			
04.05.2018	PubMed	bayes* AND analysis AND nurs*	Titel	5	5	5	pubmed_result_1
04.05.2018	PubMed	bayes* AND analysis AND social*	Titel	5	0		
04.05.2018	PubMed	bayes* AND nurs*	Titel	17	9	8	pubmed_result_2 keine deutsch-sprachige

Datum	Datenbank	Suchbegriff	Suchart	Anzahl der Treffer	Anzahl relevanter Treffer	Einschluss	Bemerkung
07.05.2018	Cochrane	bayes*	Title, Abstract, Keywords	5	0	0	

Datum	Datenbank	Suchbegriff	Suchart	Anzahl der Treffer	Anzahl relevanter Treffer	Einschluss	Bemerkung
07.05.2018	Gerolit	bayes*	Alles	10	10	6	

Datum	Datenbank	Suchbegriff	Suchart	Anzahl der Treffer	Anzahl relevanter Treffer	Einschluss	Bemerkung
08.05.2018	Cinahl	bayes*	ALL Text	2072			weitere Eingrenzung
		bayes*	Title	503			weitere Eingrenzung
		bayes* and nurs*	Title	11	10	10	
		bayes* and statistic*	Title	23	10	7	
		bayes* and analysis	Title	140			weitere Eingrenzung
		bayes* and analysis and social*	Title	1	0		
		bayes* AND analysis AND nurs*	Title	5	4	4	
		bayes* AND histori*	Title	5	0		
							keine deutsch-sprachige Suche